THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

A role for the DNA methylation system in polycomb protein-mediated gene regulation

James Reddington

Thesis presented to the University of Edinburgh for the Degree of Doctor of Philosophy
2011

Declaration

I declare that this thesis has been composed by me, and that all of the work is my own unless otherwise stated.

James Reddington
December 2011

Acknowledgements

I would first like to thank Richard for all of his supervision, support and enthusiasm during this project and also for the opportunity to be a part of his lab. I really enjoyed it and feel like I have learned a lot during this time - thank you. Thanks also to Wendy Bickmore who as my second supervisor has provided support throughout this project in addition to a second opinion. Thank you also to my graduate studies panel members, Bob Hill and Patricia Yeyati, for their time, advice and opinions during my progress meetings. All of the members of the Meehan lab, past and present, have all helped me at some stage during my time here, whether it be through practical or project advice, or just by making the lab an enjoyable environment to be in. Thank you to Richard, Colm, Donncha, Jamie, Duncan, Monika, Diana, Raffy, Meng, Alexey, Katie, John, Sara and Michael for this. The same goes for all that work in the E3 lab area and my office mates, many of whom have helped me with reagents, advice or have baked some nice cakes. Special thanks to Colm for his time spent with me figuring out the basics of R and microarray data processing, and to Shelagh for her help with the FISH technique. Thank you to collaborators from other institutes for sharing resources or performing experiments that contributed to this work, Neil Youngson, Emma Whitelaw and the labs of John Greally and Bauke Ylstra. Finally I would like to thank my wife, parents, brother and family for all of their support.

Abstract

Chromatin structure and epigenetic mechanisms play an important role in initiating and maintaining the intricate patterns of gene expression required for embryonic development. One such mechanism, DNA methylation (5mC), involves the chemical modification of cytosine bases in DNA and is implicated in maintaining patterns of transcription. However, many fundamental aspects of DNA methylation are not fully understood, including the mechanisms by which it influences transcriptional states. Recent data suggest functional links between DNA methylation and a second epigenetic mechanism that has important roles in transcriptional repression, the polycomb group (PcG) repressor system. Here, I suggest that an intact DNA methylation system is required for the repression of many PcG target genes by influencing the genomic targeting of the polycomb repressor 2 complex (PRC2) and its signature histone modification, H3K27me3 (K27me3). I demonstrate differential genomic localisation of K27me3 at gene promoter regions in hypomethylated mouse embryonic fibroblast (MEF) cells deficient for the major maintenance DNA methyltransferase, Dnmt1. Globally, *Dnmt1*$^{-/-}$ MEFs have a higher level of the K27me3 mark than controls, as assessed by western blot and immunofluorescence. I observe increased K27me3 at a relatively small number of gene promoters in *Dnmt1*$^{-/-}$ MEFs that often are associated with high levels of DNA methylation in wildtype MEFs, consistent with the notion that DNA methylation is capable of antagonising PRC2 binding at certain loci. Conversely, I show that a large number of developmentally important genes that are normally repressed and highly bound by K27me3, including classic polycomb targets, the *Hox* genes, display dramatically reduced association with K27me3 in *Dnmt1*$^{-/-}$ MEFs. Many of these genes, but not all, show reciprocal increases in promoter H3K4me3 modification and are transcriptionally de-repressed in *Dnmt1*$^{-/-}$ MEFs. I suggest that these genes are mostly associated with CpG-rich promoters with low levels of DNA methylation in wildtype cells, implying that their silencing is not dependent on the canonical role of DNA methylation. Consistent with the findings of recently published work, I suggest a working model where PRC2 binding in wildtype cells is restricted by CpG methylation. According to this model, the differential genomic location of K27me3 in hypomethylated *Dnmt1*$^{-/-}$ MEFs is explained by a redistribution of PRC2 to normally DNA methylated, unbound loci, resulting in a titration effect and coincident loss of K27me3 from normal targets. It was also apparent that certain PRC2-target genes, including the developmentally important *Hox* gene

clusters, are strongly affected in *Dnmt1*$^{-/-}$ MEFs, displaying striking loss of K27me3. As intergenic transcription has been implicated in relief from polycomb silencing and abundant intergenic transcription has been reported within *Hox* clusters, I measured RNA expression at *Hox* clusters and a small number of other PcG target genes in *Dnmt1*$^{-/-}$ MEFs using high-density tiling arrays. In *Dnmt1*-deficient MEFs, widespread increases in intergenic transcription were observed within *Hox* clusters. In addition, mapping of the elongating-polymerase-associated H3K36me3 histone modification showed widespread increases in this mark at intergenic and promoter regions in *Dnmt1*$^{-/-}$ MEFs. Increased local intergenic RNA and H3K36me3 were found to correlate with K27me3 loss for this cohort of genes. I suggest a working model where increased intergenic transcription and H3K36me3 in *Dnmt1*$^{-/-}$ MEFs leads to accelerated loss of K27me3 at certain loci, including *Hox* clusters. Taken together with recently published data, this work suggests that a major role of DNA methylation is in shaping the PRC2/K27me3 landscape. The potential implications of this putative role for DNA methylation are widespread, including our knowledge of how DNA methylation influences transcriptional regulation, and the consequence of rearranged DNA methylation patterns that are observed in many diseases including cancers.

Table of contents

List of abbreviations

This list does not include abbreviations that occur within individual gene or protein names or chemical symbols

5-Aza-dC - 5-Aza-2'-deoxycytidine
5caC - 5-Carboxylcytosine
5fC - 5-Formylcytosine
5hmC - 5'-hydroxymethylcytosine
5mC - 5'-methylcytosine
5mCpG - Methylated cytosine in the context of a CpG motif
ADD - ATRX, Dnmt3 and Dnmt3L (protein domain found in these three proteins)
AML - Acute myeloid leukemia
AP - Alkaline phosphatase
ATP - Adenosine triphosphate
AZA - 5'-deoxy-2'-azacytidine
BAC - Bacterial artificial chromosome
BAH - bromo adjacent homology
BS - Bisulfite sequencing
BSA - Bovine serum albumin
C, G, A, T - Cytosine, guanosine, adenine and thymine
cDNA - Copy DNA
CGI - CpG island
ChIP-chip - Chromatin immunoprecipitation on a microarray (chip)
CpG - cytosine base directly preceding guanosine in DNA
C-terminal - Carboxyl terminal (referring to polypeptide)
CXXC - refers to amino acid motif (cysteine, x, x, cysteine)
Cy - Cyanine
DAPI - 4',6-diamidino-2-phenylindole
DEPC - Diethylpyrocarbonate
DMEM - Dulbecco's Modified Eagle Media
DMR - Differentially methylated region
DMSO - Dimethyl sulfoxide
DNA - Deoxy-ribonucleic acid
Dnmt - DNA methyltransferase
Dnmt3 - A protein of the DNA methyltransferase 3 family

dsDNA - Double-stranded DNA

DTT - Dithiothreitol

E. coli - Escherichia coli

ECL - Electrochemiluminescence

EDTA - Ethylenediaminetetraacetic acid

ES - Embryonic stem (cells)

FDR - False discovery rate

FISH - Fluorescence in situ hybridisation

FITC - Fluorescein isothiocyanate

g - Gravitational constant

GO - Gene ontogeny

GO - Gene ontology

H - A, C or T (DNA bases)

H3K27me3 - histone H3 (H3), lysine at position 27 (K27), trimethylation (me3)

HCP - High-density CpG promoter

HDAC - Histone deacetylase

HELP - HpaII tiny fragment enrichment by ligation-mediated PCR

HELP-tag-seq - A modification of the HELP assay

HGU - Human genetics unit

HMC - Histone modifying complex

Hox - Homeotic/homeobox

HRP - Horse radish peroxidase

IAP - Intercisternal A particle

ICF - Immunodeficiency, centromere instability and facial abnormalities (a syndrome)

ICP - Intermediate-density CpG promoter

ICR - Imprinting control region

IgG - Immunoglobulin G

INP - Input (from a ChIP experiment)

IP - Immunoprecipitation

IQR - Inter-quartile range

k - kilo

LB - Luria-Bertani

LCP - Low-density CpG promoter

lincRNA - Long non-coding RNA

LINE - Long interspersed nuclear elements

M - Molar

MBD - Methyl-CpG binding domain

MD2 - MommeD2 (Modifiers of Murine Metastable Epialleles)

MeCP - Methyl-CpG-binding protein

MeDIP - Methyl-DNA-immunoprecipitation

MEF- mouse embryonic fibroblast

MLL - Mixed lineage leukemia

Mnase - Micrococcal nuclease

M-phase - mitotic phase

N-ChIP - Native chromatin immunoprecipitation

N-CoR - Nuclear co-repressor

ncRNA - Non-coding RNA

NLS - Nuclear localisation signal

nm - nanometers

N-terminal - Amino terminal (referring to polypeptide)

NTP - Nucleotide tri-phosphate

Pax - Paired box

PBS - Phospho-buffered saline

PcG - Polycomb group

PCR - Polymerase chain reaction

PGC - Primordial germ cell

PhD - Doctor of philosophy

PHD - plant homeodomain

PhoRC - Pho-repressive complex

piRNA - Piwi-interacting RNA

PMD - Partially methylated domain

PolII - RNA polymerase II

PRC1 - Polycomb repressor complex 1

PRC2 - Polycomb repressor complex 2

PR-DUB - Polycomb repressive de-ubiquitinase

pRNA - Promoter RNA

ProK - Proteinase K

PVDF - Polyvinylidene fluoride

PWWP - Refers to amino acid motif (proline, tryptophan, tryptophan, proline)

qPCR - Quantitative polymerase chain reaction

RdDM - RNA-directed DNA methylation

rDNA - Ribosomal DNA

RING - really interesting new gene (protein domain)

RIPA - Radioimmunoprecipitation

RMA - Robust multichip average

RNA - Ribo-nucleic acid

RT - Reverse transcriptase

RT-PCR - Reverse-transcription polymerase chain reaction

SAM - S-adenosyl methionine

SDS - Sodium dodecyl sulfate

SEM - Standard error of the mean

Seq - sequencing (e.g. ChIP-seq)

SINE - Short interspersed nuclear elements

S-phase - Synthesis (of DNA) phase

SRA - SET and RING associated domain

SSC - Saline-sodium citrate

ssDNA - Single-stranded DNA

SWI/SNF - SWItch/Sucrose NonFermentable

TAE - Tris-acetate-EDTA

TBE - Tris-borate-EDTA

tDMR - Tissue-specific differentially methylated region

TE - Tris EDTA

TET - Ten-eleven translocation (protein family)

TF - Transcription factor

TRD - Transcription repression domain

tRNA - Transfer RNA

TrxG - Trithorax group

TSA - Trichostatin A

TSS - Transcription start site

TxRd - Texas red

UCSC - University of Santa Cruz, California, USA

UTR - Un-translated region

UV - Ultra-violet

WT - Wildtype

Xa - Active X chromosome

X-ChIP - Cross-linked chromatin immunoprecipitation

XCI - X-chromosome inactivation

Xi - Inactive X chromosome

αKG - α-keto gluterate

Chapter 1 - Introduction

A major challenge of molecular biology is to decipher the molecular organisation of eukaryotic genomes and to establish the role that it plays in coordinating gene expression. In eukaryotes, nuclear DNA is complexed with small basic scaffold proteins and other protein components in a structure referred to as chromatin. Chromatin organisation occurs at multiple levels within the nucleus and is dynamic allowing it to play a crucial role in the regulation of DNA-dependent processes such as transcription (reviewed in Li *et al.,* 2007). Considerable effort has been applied to the study of chromatin organisation in recent years and while we are beginning to understand some of the processes involved, many fundamental questions remain unanswered. Of particular interest are the mechanisms by which chromatin organisation contributes to the complex regulation of gene transcription necessary for multi-cellular life. Mammalian development involves the specification and maintenance of over 200 cell types, via multiple developmental transitions, that originate from one original totipotent cell (Hochedlinger and Plath, 2009). During development, transcriptional changes are initiated by the action of transcription factor networks. Changes in chromatin structure occur in response to this process and can contribute to the maintenance of transcription states (Li *et al.,* 2007). The term 'epigenetics' is often used to describe processes involved in regulating genome activity that are independent of changes in DNA sequence (Bird, 2007). However, the meaning of this term is often confounded, perhaps owing to both the ambiguity of this definition and the existence of multiple definitions. The term 'epigenetic' was more precisely defined as a mechanism involving; 'the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states' (Bird, 2007), and this latter definition is assumed to be the definition of epigenetic mechanisms throughout this thesis. Eukaryotic cells could be said to have a memory, imparted by epigenetic mechanisms, that contributes to their maintenance of gene transcription patterns through time and/or cell divisions. A detailed comprehension of epigenetic mechanisms is therefore key to understanding how patterns of gene expression are maintained in normal development and how these processes malfunction in human disease. In this chapter I will begin by briefly discussing examples of how epigenetic mechanisms are involved in gene regulation, followed by a more detailed review of our knowledge of two such mechanisms, DNA methylation and the polycomb-group (PcG) repressor system, which are the subjects of this thesis.

1.1. Epigenetic regulation of transcription

In humans, approximately two metres of DNA is packed into the nucleus (Alberts *et al.*, 2002). It is believed that this level of compaction is achieved by multiple levels of chromatin organisation. This organisation is currently better understood at the lower levels (i.e. the nucleosome) while the higher levels of chromatin organisation remain relatively obscure. This thesis focuses on mechanisms that are thought to act mainly at the lower levels of chromatin organisation and therefore higher levels of organisation will only be touched upon. The primary level of chromatin compaction is achieved by wrapping DNA around a protein scaffold formed, in most eukaryotes, by histone proteins. This fundamental repeating unit of chromatin is the nucleosome, DNA associated with an octomer of histone proteins (Burlingame *et al.,* 1985; Luger *et al.,* 1997). The core particle is the name given to 146bp of DNA wrapped ~1.7 times around the histone octomer (Burlingame *et al.*, 1985; Luger *et al.,* 1997; van Holde and Zlatanova, 1999). The histone octomer itself is composed of two molecules of each of the four types of histone protein. In most eukaryotes, nucleosomes are generally composed of the core histones, H2A, H2B, H3 and H4 (Burlingame *et al.*, 1985; Luger *et al.,* 1997). The core histones each have a distinct C-terminal, structured, globular domain and a flexible, unstructured N-terminal tail which protrudes from the nucleosome (Schroth *et al.*, 1990; Luger *et al.,* 1997). In addition to the nucleosomal histones, linker histones form part of chromatin and are thought to influence its structure at a higher level. Mammalian genomes encode many linker histone genes that code for proteins of the H1 and H5 histone families that are structurally distinct from nucleosomal histones (reviewed in McBryant *et al.*, 2010).

Importantly, chromatin organisation is dynamic and variable across different regions of the genome and between cell types. Many mechanisms have been proposed to influence gene expression through chromatin dynamics. A very brief overview of some of the best studied mechanisms is given here.

Firstly, nucleosomes can be mobile on the DNA template meaning that nucleosome positioning can be variable (Pennings *et al.*, 1991). RNA polymerase activity is thought to be inhibited by packaging DNA into nucleosomes (Huang and Bonner, 1962; Allfrey *et al.*, 1963) and some DNA binding factors are thought to be inhibited by the positioning of a nucleosome over their recognition site (Almer *et al.*, 1986; reviewed in Cairns, 2009), leading to the opinion that nucleosome positioning can influence transcription. In support of this idea, gene promoters and enhancer regions often contain a nucleosome depleted region

that can be constitutive and mediated by DNA sequence elements or variable and dependent on nucleosome remodelling factors (Struhl, 1985; reviewed in Cairns, 2009). Indeed, proteins that use ATP to reposition or eject nucleosomes, nucleosome remodelers, are often required to establish and maintain patterns of transcription (Whitehouse *et al.*, 1999; Lorch *et al.*, 1999; reviewed in Cairns, 2009). Also, a characteristic pattern of nucleosome positioning exists downstream of transcription start sites that is established, at least in part, by active ATP-dependent mechanisms, not transcription itself, highlighting the importance of nucleosome positioning out with of promoter regions (Zhang *et al.*, 2011a).

Secondly, variants of histone proteins can be incorporated into nucleosomes and affect chromatin structure by possessing different molecular properties from their core histone analogues. All of the core histones, with the exception of H4, have known variants in mammals (reviewed in Banaszynski *et al.*, 2010). Variants can contain minor sequence differences, such as the variant of H3, H3.3, or significant structural differences, such as macroH2A (Pehrson and Fried, 1992; reviewed in Banaszynski *et al.*, 2010). Histone variants appear to play crucial roles in chromatin structure and transcriptional regulation. For example, the H2A variant, H2A.Z, is essential for mouse embryogenesis (Faast *et al.*, 2001). H2A.Z is abundant in chromatin and was found to be present primarily at promoter regions in the yeast *Saccharomyces cerevisiae* (Zhang *et al.*, 2005). H2A.Z has many proposed functions including promoting gene activation by increasing access to DNA (Zhang *et al.*, 2005) and, conversely, in heterochromatin formation (Rangasamy *et al.*, 2003).

Thirdly, linker histones are thought to influence chromatin structure. *In vitro*, it is well established that linker histones affect chromatin structure by promoting the conversion of a heterogeneous population of nucleosomal arrays into a 30nm fibre, an enigmatic higher order chromatin structure (Thoma *et al.*, 1979). This chromatin compaction appears to be achieved by the ability of linker histones to bind inter-nucleosomal DNA, termed linker DNA, at the nucleosomal DNA entry/exit sites (Buckle *et al.*, 1992; Thomas *et al.*, 1992; reviewed in McBryant *et al.*, 2010). The relative amount of linker histones in chromatin has been linked to transcriptional regulation as actively transcribed chromatin regions are thought to be depleted of linker histone relative to inactive regions (Woodcock *et al.*, 2006; reviewed in McBryant *et al.*, 2010). However, the role of linker histones in gene expression remains unclear as embryonic stem cells with depleted histone H1 showed only a small number of gene expression changes despite dramatic changes in chromatin structure (Fan *et al.*, 2005).

Fourthly, histones are subject to a wealth of post-translational modifications, most of which occur within the unstructured N-terminal tails, many of which are associated with

transcriptional states (Cao *et al.*, 2002; reviewed in Kouzarides, 2007). Currently, there are ~100 different histone modifications identified, many of which are proposed to influence the presence of others, highlighting the complexity of this level of chromatin structure, sometimes referred to as the 'histone code' (Turner, 2000; Jenuwein and Allis, 2001; Tan *et al.*, 2011). Among the well studied histone modifications with respect to gene regulation are methylation and acetylation of lysine residues. Modifications of histones are thought to influence chromatin structure in two ways; (1) by affecting nucleosome-nucleosome or nucleosome-DNA interactions directly, or, (2) by influencing the binding of proteins to chromatin (reviewed in Kouzarides, 2007). Acetylation of lysines was predicted to cause local decompaction of chromatin by neutralising the basic charge of the lysine residue, hence affecting its interactions. Indeed, acetylation of H4K16 inhibits both the formation of the 30nm fibre and the binding of an ATP-dependent chromatin remodeler, ACF, *in vitro*, fulfilling both of the proposed functions of a histone modification (Shogren-Knaak *et al.*, 2006). A conserved protein domain, the bromodomain, has been identified that binds to acetylated lysine and is present in many proteins with roles in transcriptional regulation (Tamkun *et al.*, 1992). In addition, mapping of histone acetylation has shown that it is generally a marker of active gene promoters (Kurdistani *et al.*, 2004; reviewed in Choi and Howe, 2009) contributing to the idea that histone acetylation is tightly coupled with the process of transcription initiation. Recently, histone acetylation at the lysine 27 (K27) of the histone H3 tail has been shown to be present at enhancer elements in mammals, particularly at 'active' enhancers, expanding our view of the roles of histone acetylation outside of annotated gene promoters (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011). It is interesting to speculate that histone acetylation at active enhancers serves a similar purpose as the proposed role at gene promoters, by allowing for a more open chromatin structure that is more compatible with protein-DNA interactions. In contrast to lysine acetylation, lysine methylation has been associated with roles in both gene activation and repression depending on the particular lysine residue that is modified and the location of the modification relative to genes. For example, di- and tri-methylation of lysine at position 9 of histone H3 (H3K9me2 and H3K9me3) are associated with transcriptional repression when present at the promoter region of genes (reviewed in Kouzarides, 2007). H3K9me3 is required for the formation of large nuclear heterochromatin compartments, such as pericentromeric heterochromatin (Schotta *et al.*, 2004), and is also linked to the repression of genes in euchromatic regions of the genome (Tachibana *et al.*, 2002). Another well studied form of lysine methylation is H3K27me3 which is deposited by polycomb group proteins. This modification will be discussed in detail in a later section of this chapter.

In addition to post-translational modification of histone proteins, chromatin can be modified on DNA molecules, in the form of DNA methylation, and this can contribute to the regulation of transcription. The following section summarises our knowledge of the DNA methylation system and its role in gene regulation.

1.2. DNA methylation

1.2.1. The mark

In mammals, the cytosine nucleotide of DNA can exist in at least three forms including (but not exclusive to) un-modified cytosine, methyl-cytosine and hydroxymethyl-cytosine, representing the only known modifications of DNA to date (reviewed in Bird, 2011). Methyl-cytosine (5mC) is formed by the addition of a methyl group to the 5' carbon of the pyrimidine ring of cytosine, a reaction catalysed by DNA methyltransferases (Fig. 1.1). Hydroxymethyl-cytosine (5hmC) is catalysed by the hydroxylation of methyl-cytosine by enzymes of the Tet (Ten Eleven Translocation) family (Fig. 1.1).

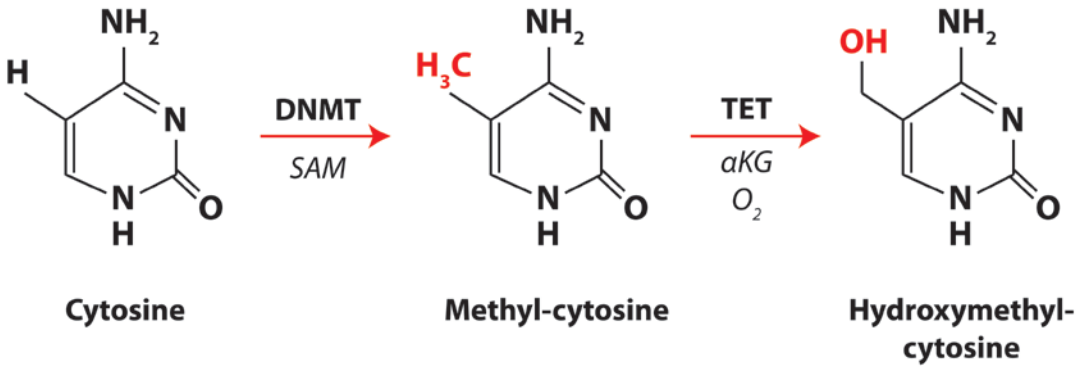


Figure 1.1 - The three major forms of the cytosine base. Cytosine can be converted to methyl-cytosine by DNA methyltransferases (DNMTs) with co-factor S-adenosyl methionine (SAM). Methyl-cytosine can be hydroxylated to form hydroxymethyl-cytosine by Ten-eleven-translocation (TET) family proteins in the presence of α-keto gluterate (αKG) and oxygen. Other modified forms of cytosine have been identified but are present at very low levels where identified so far.

In animals, 5mC, herein referred to as DNA methylation, exists mainly where cytosine is adjacent to guanosine in a 5'-CG-3' context, although it has been reported in other sequence contexts (discussed in 1.2.4.7). In vertebrates, it is estimated that 70-80% of cytosines in the CG context are methylated (5mC) with the exception of CpG islands, stretches of CpG- and GC-rich DNA, leading to an overall bimodal distribution of the mark (reviewed in Ndlovo *et al.*, 2011). Importantly, the methyl group of 5mC lies within the

major groove of B-form DNA and so does not change base pairing properties of the DNA template (see for example Ohki *et al.*, 2001).

DNA methylation is proposed to play crucial roles in the regulation of transcription and the formation of epigenetic memory. This, in addition to our knowledge of its distribution, mechanisms of action and dynamics are discussed in the following sections.

1.2.2. The mediators - DNA methyltransferases

The enzymes responsible for catalysing methylation (5mC) of cytosine are DNA methyltransferases (DNMTs). The reaction mechanism of the DNMTs requires a methyl donor, S-adenosyl methionine (SAM) and is thought to occur as follows (O'Gara *et al.*, 1996; reviewed in Goll and Bestor, 2005). A covalent bond is formed between a conserved cysteine residue of the methyltransferase and the C6 carbon of the cytosine ring and protonation of the N3 position results in a reactive intermediate which attacks the methyl group of SAM leading to methyl transfer to the C5 position of cytosine (Santi *et al.*, 1983; Chen *et al.*, 1991). Subsequently, the enzyme is released through β elimination (O'Gara *et al.*, 1996; reviewed in Goll and Bestor, 2005). This reaction depends on six amino acid motifs, that are well conserved in all 5'-cytosine methyltransferases, that are involved in the processes described above (reviewed in Goll and Bestor, 2005).

Here I discuss our knowledge of the different families of DNMTs with a strong emphasis on the vertebrate proteins. Unless otherwise stated, it should be assumed that I am referring to vertebrate proteins.

1.2.2.1. Dnmt1 and maintenance methylation

It is thought that DNA methylation patterns, with the exception of specific regions and in response to certain developmental signals, are relatively stable. It was shown that methylation, in the form of a 5mCpG dinucleotide, is stably maintained in mouse cells across cell divisions (Stein *et al.*, 1982). This property of DNA methylation is fundamental for a role in epigenetic memory where patterns of transcription must be maintained. This role is fulfilled by an exquisite mechanism for maintenance methylation dependent on the so called maintenance methyltransferase, Dnmt1, and a co-factor, Uhrf1. During S-phase of the cell cycle, DNA is replicated in a semi-conservative manner producing two double-stranded DNA molecules, each composed of one of the parent strands and a nascent strand (Meselson

and Stahl, 1958). This means that the DNA methylation pattern from the parent DNA is still present on the parent strand and is now paired with an unmethylated nascent strand. We now know that the process of maintenance methylation copies the pattern of methylation from the parent strand to the nascent strand leading to propagation of DNA methylation patterns across cell divisions. This process is achieved by the affinity of Dnmt1 for hemi-methylated DNA (Yoder *et al.,* 1997) and the action of a maintenance methylation co-factor, Uhrf1. Uhrf1 (also known as Np95 and ICBP90) was shown to be required for normal levels of DNA methylation and to interact with the Dnmt1 protein. Uhrf1 contains a methyl-CpG binding SRA (SET and RING associated) domain with a strong preference for hemi-methylated CpG motifs leading to the idea that it may recruit Dnmt1 to hemi-methylated CpGs (Bostick *et al.,* 2007; Sharif *et al.,* 2007; Avvakumov *et al.,* 2008). In support of this model, Uhrf1 and Dnmt1 co-localise at replication foci during S-phase, Uhrf1 is required for full association of Dnmt1 with replication foci (Bostick *et al.,* 2007; Sharif *et al.,* 2007) and *Uhrf1*$^{-/-}$ mice show a similar embryonic lethality phenotype to *Dnmt1*-mutants (Sharif *et al.,* 2007). The SRA domain of Uhrf1 binds to hemi-methylated DNA by a base flipping mechanism where the methylated base is rotated out into a protein binding pocket of Uhrf1. Subsequently, the unmethylated cytosine on the other strand is flipped out, suggesting that it may be presented in this way to Dnmt1 (Arita *et al.,* 2008). In addition to Uhrf1, Dnmt1 is targeted to replication foci by an interaction with PCNA (proliferating cell nuclear antigen) which modestly increases the fecundity of maintenance methylation by about two-fold (Spada *et al.,* 2007; Schermelleh *et al.,* 2007)(summarised in fig. 1.2).

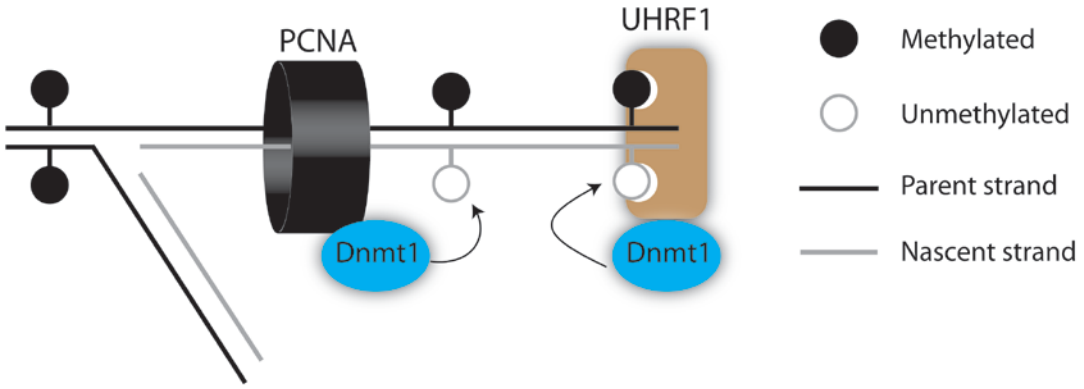


Figure 1.2 - Model for maintenance methylation. DNA methylation patterns are maintained across cell divisions by Dnmt1. Dnmt1 is targeted to replication foci and hemi-methylated CpGs by interactions with PCNA and Uhrf1.

The above model is proposed to account for the vast majority of maintenance methylation but it is likely that other factors are involved to complete the process. For example, it has been suggested that Dnmt3 proteins may 'top up' methylation that is missed by an imperfect Dnmt1-based mechanism (Jones and Liang, 2009).

The structure of the Dnmt1 protein is bipartite, containing a C-terminal catalytic domain and an N-terminal regulatory domain separated by an unstructured linker (Goll and Bestor, 2005; Turek-Plewa and Jagodzinski, 2005)(Fig. 1.3). The catalytic domain contains the conserved motifs characteristic of a 5'-cytosine methyltransferase. The N-terminal portion has many conserved and functional domains. At the N-terminus of the protein is a domain responsible for binding to DMAP (Dnmt1 associated protein 1), a protein implicated in transcriptional repression (Rountree *et al.*, 2000). DMAP1 is recruited to replication foci by Dnmt1 and this complex is joined by HDAC2 (histone deacetylase 2) late in S-phase (Rountree *et al.*, 2000). However, the functional significance of the Dnmt1-DMAP1 interaction is unclear as mice that express only Dnmt1 that lacks the DMAP-interaction domain are phenotypically normal (Ding and Chaillet, 2002). The N-terminal portion of the protein also contains a domain responsible for nuclear localisation and a replication foci targeting domain that is alone capable of targeting Dnmt1 to PCNA foci during S-phase (Leonhardt *et al.*, 1992). Interestingly, Dnmt1 forms foci at heterochromatin in G2 and M-phases of the cell cycle and is continuously loaded onto chromatin during this time. This localisation is dependent on a different domain than the targeting domain described above and is independent of heterochromatin components HP1 and H3K9me3 (Easwaran *et al.*, 2004). The function of this localisation is currently unclear but it suggests that Dnmt1 may function outside of S-phase. The two BAH (bromo adjacent homology) domains are of

unknown function but are present in other chromatin-associated proteins and have been suggested to be involved in protein-protein interactions (Goll and Bestor, 2005).

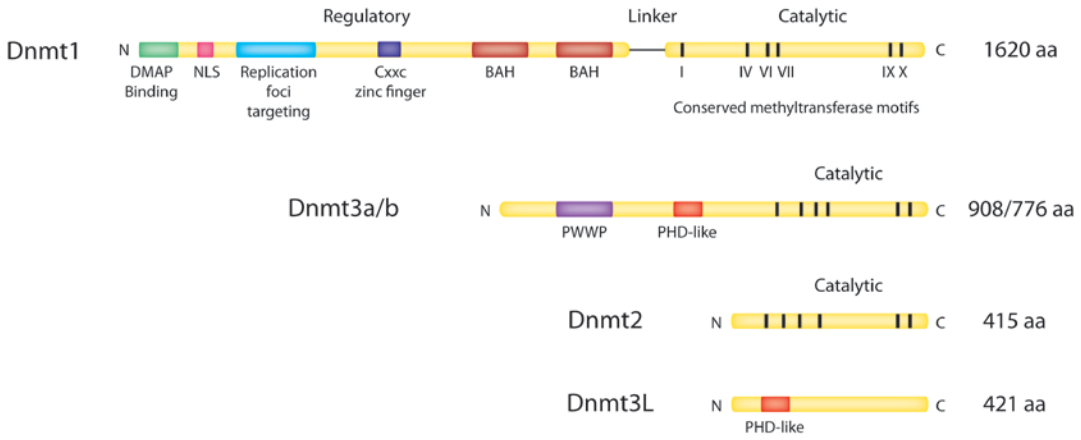


Figure 1.3 - Domain structure of the mammalian DNA methyltransferase family proteins. NLS: nuclear localisation signal. BAH: bromo adjacent homology. CXXC and PWWP refer to amino acid motifs present in each domain. PHD: plant homeodomain. Adapted from Goll and Bestor (2005) and Turek-Plewa and Jagodzinski (2005).

Dnmt1 contains a CXXC zinc finger domain, a domain also found in Cfp1 and Mll1 proteins, that is associated with DNA binding. The presence of this domain was intriguing as it is known to bind unmethylated CpG motifs in other proteins (Thomson *et al.*, 2010). Recent structural data has shed light on the possible function of this domain in maintenance methylation where it was suggested to play an autoinhibitory role (Song *et al*,. 2011). Specifically, when bound to an unmethylated CpG, the CXXC domain places the CXXC-BAH linker between the DNA and active site of the enzyme preventing *de novo* methylation. This idea is supported by an increased *de novo* methyltransferase activity of Dnmt1 either lacking, or with mutation in, the CXXC domain (Song *et al*,. 2011). Further structural studies suggest that methylation by Dnmt1 is a multi-step and highly regulated process (Takeshita *et al.*, 2011). These data begin to provide an explanation for how Dnmt1 methyltransferase activity is regulated and how it is intrinsically inhibited for *de novo* methyltransferase activity, a potentially harmful property for a maintenance methylase.

Dnmt1 is widely expressed and particularly highly abundant in dividing cells (Goll and Bestor, 2005). Functional studies of *Dnmt1* and its orthologs have shown that they are essential for many biological processes in a variety of organisms including mouse, human, frog, fish, plant, invertebrate and fungal species (discussed in section 1.2.5.1). Although most of the phenotypes of *Dnmt1* mutation are commonly attributed to decreases in DNA

methylation, in most cases it cannot be formally excluded that DNA methylation independent roles of DNA methyltransferases contribute to the defects. Indeed, roles for Dnmt1 outside of methylation have been proposed but it remains unclear how important they are as a catalytic *Dnmt1* mutation produces the same ES cell differentiation and embryonic lethality phenotypes as null mutation (Damelin and Bestor, 2007; Takebayashi *et al.,* 2007). However, it has emerged that the catalytic and non-catalytic roles of Dnmt1 are difficult to separate as proper localisation of Dnmt1 during S-phase requires pre-existing methylation (Takebayashi *et al.,* 2007). Suggestive evidence for methylation independent roles of Dnmt1 exists. Suggested interactors of Dnmt1 include histone methyltransferases, histone deacetylases, transcription factors, chromatin binding proteins, methyl-binding proteins and putative chromatin remodelers (discussed in section 1.2.6). In *Xenopus laevis*, catalytically inactive human DNMT1 can rescue some of the defects associated with xDnmt1 depletion (Dunican *et al.,* 2008). DNMT1 is also suggested to modulate expression of E-cadherin in human cancer cells, an effect that is rescued by the presence of catalytically inactive protein (Espada *et al.,* 2011). Roles for Dnmt1 in genomic stability have also been proposed based on observed functions in the DNA damage response and apoptosis. For example, Dnmt1 interacts with repair proteins, is recruited to laser-induced DNA double strand breaks and is required for their timely repair (Ha *et al.,* 2011). Currently, a role for Dnmt1 out with of DNA methylation is considered atypical and further investigation in this area is needed.

The stability and activity of Dnmt1 must be tightly regulated to ensure its proper function. In this regard, an interesting interplay between post-translational modifications of Dnmt1 and its interacting proteins has emerged (reviewed in Denis *et al.,* 2011). In addition to its direct role in maintenance methylation, Uhrf1 contains ubiquitin-ligase activity and may control the stability of Dnmt1 through ubiquitination leading to proteolysis (Qin *et al.,* 2011). The two proteins are bridged by Usp1, a de-ubiquitinating enzyme, in a trimeric complex where the result is thought to be fine tuning of Dnmt1 protein levels (Qin *et al.,* 2011). Usp7 has been shown to increase both the maintenance and *de novo* methylation activity of Dnmt1 *in vitro* suggesting that it may participate at multiple levels in Dnmt1 regulation (Felle *et al.,* 2011a). In addition to ubiquitination, Dnmt1 may be regulated by lysine methylation and serine phosphorylation. It is thought that lysine methylation by Set7/9 destabilises Dnmt1 protein and that a balance is struck with Lsd1 (lysine specific demethylase 1) to obtain appropriate Dnmt1 levels (Wang *et al.,* 2009a). Also a phosphorylation-methylation switch has been suggested to control Dnmt1 protein stability. Specifically, phosphorylation of Ser143 of Dnmt1 by AKT1 is mutually exclusive with methylation of the adjacent Set7/9 target Lys142 residue. This means that Ser143

phosphorylated Dnmt1 is more stable. Consistent with this idea, phosphorylation of Dnmt1 peaks just before S-phase when abundant Dnmt1 protein is required for maintenance methylation (Esteve *et al.,* 2010). Heterozygous mutations in DNMT1 have been associated with hereditary sensory neuropathy with dementia and hearing loss, a dominant neuro-degenerative disorder (Klein *et al.,* 2011). The mutations consisted of one to three amino acid substitutions which lead to misfolding, instability and reduced methyltransferase activity of the DNMT1 protein.

1.2.2.2. Dnmt3 proteins and de novo methylation

The establishment of DNA methylation on previously unmethylated cytosines, *de novo* methylation, is performed primarily by two homologous Dnmt3 proteins, Dnmt3a and Dnmt3b. These proteins are similar to Dnmt1 in that they comprise an N-terminal domain implicated in protein-protein interactions and chromatin binding and a C-terminal catalytic domain (Fig. 1.3). Their N-terminal portions contain PWWP domains, named after the amino acid motif that it contains, and a cysteine-rich PHD-like (plant homeodomain-like) domain. Both are thought to be involved in chromatin binding and recognition of specific histone modifications (reviewed in Denis *et al.,* 2011). The PWWP domain of Dnmt3b is required for association of the protein with, and methylation of, pericentric heterochromatin implicating it in chromatin binding (Chen *et al.* 2004; Ge *et al.,* 2004). The PHD-like domain (sometimes referred to as the ADD domain - ATRX, Dnmt3 and Dnmt3L domain) is not absolutely required for this localisation but may facilitate it (Chen *et al.* 2004). Both the PWWP and PHD-like domains of Dnmt3a are also implicated in targeting *de novo* methylation to chromatin containing certain histone modifications (discussed in section 1.2.4.11).

Dnmt3 proteins are expressed in a range of adult tissues but are less abundant than Dnmt1 (Goll and Bestor, 2005). Dnmt3b is expressed highly in ES cells, early development and in germ cells while its expression is very low in most somatic cells apart from thyroid and bone marrow (Xie *et al.,* 1999). Dnmt3a is thought to be expressed at low levels ubiquitously but is abundant in ES cells (Okano *et al.,* 1998a; Xie *et al.,* 1999; Chen *et al.,* 2002). Functional studies of Dnmt3a and Dnmt3b suggest that they have non-overlapping functions in development. Although null mutation of each gene in mouse is incompatible with viability, the phenotypes and stage of lethality differ considerably (Okano *et al.,* 1999a)(discussed further in 1.2.5.1). This suggests that each enzyme has specific targets. Indeed, Dnmt3b is required for methylation of pericentric repeats which contributes to

chromosomal stability, and Dnmt3a cannot substitute for the absence of Dnmt3b in this respect (Okano *et al.*, 1999a). Conversely, Dnmt3a, but not Dnmt3b, is required for establishing most imprinting during male and female germ cell development (Kaneda *et al.*, 2004).

Early *in vitro* studies showed that Dnmt3 proteins have no preference for hemi-methylated over unmethylated CpGs suggesting that they may perform as *de novo* methyltransferases (Okano *et al.*, 1998a). The activity of methyl transfer *in vitro* by Dnmt3s is low, consistent with the idea that other factors stimulate *de novo* methylation *in vivo* (Okano *et al.*, 1998a). Dnmt3a may also methylate cytosine outside of the CpG context, primarily at CpA and CpT dinucleotides, although its major target is CpG (Ramsahoye *et al.*, 2000)(discussed further in 1.2.4.7). It has been suggested that Dnmt3 proteins have a DNA sequence preference based on the nucleotides flanking the central CpG dinucleotide as certain sequences were found to be preferentially methylated *in vivo* and *in vitro* (Handa and Jeltsch, 2005; Zhang *et al.*, 2009; Wienholz *et al.*, 2010). It has been reported that Dnmt3b has a higher methyltransferase activity on nucleosomal DNA than Dnmt3a (Shen *et al.*, 2010). Interestingly, this appears to be a mammal-specific adaption of Dnmt3b as non-mammalian orthologous proteins do not show increased activity on nucleosomal DNA. This variation may be crucial for proper methylation of relatively compact nucleosomal repetitive DNA in mammals, and has thus been linked to the rapid expansion of such sequences in mammalian genomes (Shen *et al.*, 2010).

Dnmt3a and Dnmt3b proteins, in contrast to Dnmt1, are tightly associated with chromatin containing methylated DNA, a property that depends on the presence of an intact N-terminal region (Jeong *et al.*, 2009). In fact, it has been suggested that free Dnmt3 proteins are subject to degradation, therefore methylated DNA-containing nucleosomes have a stabilising effect on these proteins (Sharma *et al.*, 2011). Dnmt3b preferentially associates with condensed, histone H1 containing chromatin *in vivo* but does not show a clear preference for chromatin enriched in any repression-associated histone marks such as H3K9me3 (Kashiwagi *et al.*, 2010). There is, however, a modest depletion of activation-associated marks, H3K4me3 and H4K16ac, in Dnmt3b-bound chromatin. This may be a cause or consequence of Dnmt3b binding as Dnmt3b associates with histone deacetylases (Kashiwagi *et al.*, 2010).

Dnmt3 proteins are thought to interact with other Dnmt3 proteins, including non-catalytic isoforms, to form homo- and heterodimers (reviewed in Cheng and Blumenthal, 2008). The dimerisation of Dnmt3 proteins is thought to influence various properties of the enzymes (van Emburgh *et al.*, 2011). Dnmt3L is a catalytically inactive DNMT-like protein

with roles in regulation of Dnmt3 proteins (discussed in 1.2.2.4). Models of a tetrameric complex consisting of two Dnmt3a-Dnmt3L heterodimers bound to DNA suggest that they could methylate two CpGs separated by one helical turn (Jia *et al.,* 2007; Cheng and Blumenthal, 2008). In agreement with this model a periodicity of 9 bp has been observed in highly methylated regions of human chromosome 21 (Zhang *et al.,* 2009) and of imprinted control regions (Jia *et al.,* 2007). Other suggested interactions of Dnmt3 proteins include Dnmt1 and chromatin modifying proteins such as HDACs and hSNF2H suggesting that, like Dnmt1, non-catalytic roles of Dnmt3 proteins may be important for their function (Kim *et al.,* 2002; Geiman *et al.,* 2004).

Mutations in human *DNMT3* genes are associated with pathological conditions. Mutations in *DNMT3B* cause the rare autosomal recessive disorder ICF (immunodeficiency, centromere instability and facial abnormalities) syndrome, a disorder characterised by the three main symptoms from which it is named (Xu *et al.,* 1999). Two thirds of ICF patients have *DNMT3B* mutations (Jin *et al.,* 2008). A major contributor to these phenotypes is thought to be severe hypomethylation of pericentric repeats leading to chromosomal abnormalities, a problem that manifests itself most strongly in cells of the immune system (Xu *et al.,* 1999; reviewed in Goll and Bestor, 2005). It is not clear if hypomethylation of pericentric regions is the only contributor to the disorder as defects in regulation of developmental genes and disorganisation of chromosome territories have also been observed in patient cell lines (Matarazzo *et al.,* 2007; Jin *et al.,* 2008). ICF-causing missense mutations throughout the protein have been identified, including the PWWP domain and many throughout the catalytic domain (Xu *et al.,* 1999). ICF mutations are thought to affect various biochemical properties of DNMT3B including SAM binding and utilisation, DNA binding, and homodimerisation (Moarefi and Chedin, 2011). Recently, mutations in human *DNMT3A* have been found in patients with certain subtypes of leukemia. In one study, 22.1% of acute myeloid leukemia (AML) patients were found to carry a non-silent mutation in *DNMT3A* and presence of *DNMT3A* mutation correlated with poor prognosis (Fey *et al.,* 2010). In another study, 20.5% of patients with acute monocytic leukemia (a subtype of AML, AML-M5) were found to carry *DNMT3A* mutations (Yan *et al.,* 2011). Lower levels of DNA methylation were observed at candidate regions in *DNMT3A*-mutants relative to controls correlating with increased transcription of certain *Hox* genes thought to be important in haematopoiesis (Yan *et al.,* 2011). However, the cause of leukemiagenesis in *DNMT3A*-mutation carriers is currently unclear but the striking proportion of *DNMT3A* mutations found in these studies points towards a major contribution of DNMT3A to normal haematopoiesis.

1.2.2.3. Dnmt2

A third family of DNMT enzymes exists, known as Dnmt2 proteins, and they have a controversial history in terms of their biological function. Dnmt2 consists solely of a catalytic domain and contains all the motifs required for catalytic activity on cytosine, making it an exciting candidate for a DNA methyltransferase when it was originally cloned (Fig. 1.3)(Yoder and Bestor, 1998; Okano *et al.,* 1998b). In fact, Dnmt2 is the most widely conserved of the DNMT proteins with homologs in *Drosophila melanogaster* and *Schizosaccharomyces pombe* (Hermann *et al.*, 2003). However, it was noticed that Dnmt2 is more similar to a fission yeast methyltransferase than to Dnmt1, fission yeast having no DNA methylation (Yoder and Bestor, 1998). Indeed, early functional studies suggested that Dnmt2 is not a major DNA methyltransferase as it was unable to methylate DNA *in vitro* and deletion of its catalytic domain in mammalian cells had no effect on methylation of newly integrated viral DNA (Okano *et al.*, 1998b). A few years later, weak residual cytosine methyltransferase activity of Dnmt2 was shown *in vitro* and Dnmt2 depletion during *Drosophila* development abolished detectable low-level DNA methylation (Hermann *et al.,* 2003; Kunert *et al.*, 2003). A breakthrough came when it was discovered that Dnmt2 is present in the cytoplasm and responsible for methylation of the tRNA for aspartic acid, a property that is now generally considered to be its major conserved function (Goll *et al.,* 2006). Dnmt2 has since been shown to methylate other tRNAs, a process responsible for protecting them against stress-induced cleavage (Schaefer *et al.,* 2010a). In keeping with this idea, Dnmt2 depletion in zebrafish results in developmental defects but only its cytoplasmic localisation is required (Rai *et al.,* 2007). However, more recently, Dnmt2 protein in *Drosophila* was shown to associate with the nuclear matrix and have access to DNA during mitosis (Schaefer *et al.*, 2008). In addition, evidence was presented that *Drosophila* Dnmt2 is required for transposon silencing and telomere integrity. This report suggested that Dnmt2 methylates DNA of retrotransposons in early development and that this was essential for heterochromatinisation and silencing of a particular type of repeat (Phalke *et al.,* 2009). Another report supported this idea by suggesting that DNA methylation of transposons does indeed occur during *Drosophila* development via a mechanism that involves Dnmt2 recruitment by SETDB1 (Gou *et al.,* 2010). However, DNA methylation of *Invader4* elements, the transposon class studied in Phalke *et al*, and the role of Dnmt2 in DNA methylation has been strongly questioned (Schaefer *et al.,* 2010b). Here, the authors suggest that there is no convincing evidence for DNA methylation at these

elements in *Drosophila* and therefore the Dnmt2 protein, in regard to its putative function in DNA methylation, remains an enigma.

1.2.2.4. Co-factors of the DNA methylation machinery

In addition to catalytically active DNA methyltransferases, many protein co-factors are required to establish and maintain DNA methylation patterns. Here I discuss a few of the best studied co-factors of DNA methylation, although more have been suggested.

One such protein is Dnmt3L, a mammal-specific DNMT-like protein that lacks any catalytic activity (Fig. 1.3)(Hata *et al.*, 2002). Dnmt3L resembles a portion of Dnmt3 proteins and contains a PHD-like domain similar to that of Dnmt3 proteins (Fig. 1.3). Dnmt3L is expressed in germ cells and the early embryo at periods where widespread *de novo* methylation is taking place (Bourc'his *et al.*, 2001). The role of Dnmt3L appears to be in the regulation and targeting of Dnmt3 proteins for *de novo* methylation (Hata *et al.*, 2002). Mice lacking *Dnmt3l* develop normally but males are infertile due to azoospermia (Bourc'his *et al.*, 2001; Hata *et al.*, 2002). Female mice lacking Dnmt3L are fertile but their heterozygous offspring die in mid-gestation, even when transplanted into a surrogate uterus (Bourc'his *et al.*, 2001; Hata *et al.*, 2002). This suggested that Dnmt3L may be required for establishing imprints, DNA methylation marks laid down in germ cells that will control parent of origin specific gene expression in the offspring. Indeed, this appears to be the case as oocytes from *Dnmt3l*$^{-/-}$ mice show severely reduced methylation at imprint control regions and embryos from *Dnmt3l*$^{-/-}$ oocytes have biallelic expression of normally imprinted genes (Bourc'his *et al.*, 2001). Also, Dnmt3L was shown to be required for DNA methylation and silencing of different retrotransposon classes during spermatogenesis and *Dnmt3l*$^{-/-}$ testes undergo meiotic defects (Bourc'his *et al.*, 2004). Overall, the *Dnmt3l*$^{-/-}$ mice phenocopy conditional knock out of *Dnmt3a* in germ cells (Kandeda *et al.*, 2004). It is now thought that Dnmt3L is required to stimulate and target Dnmt3 enzymes during specific waves of *de novo* methylation during development. In particular, it is believed that Dnmt3L targets *de novo* methyltransferases Dnmt3b and Dnmt3a2 to chromatin containing H3K4me0 (non-methylated K4). This is thought to be achieved through an interaction between the PHD-like domain of Dnmt3L and the N-terminal tail of histone H3, an interaction that is inhibited by K4me (Ooi *et al.*, 2007). In addition to targeting, Dnmt3L stimulates Dnmt3a activity by increasing its binding to SAM, the methyl donor (Chedin *et al.*, 2002; Kareta *et al.*, 2006). This activity may be caused by a Dnmt3L-mediated conformational change in the active site

of Dnmt3a and/or by the interaction surfaces formed which stabilise Dnmt3a-SAM interactions (reviewed in Cheng *et al.,* 2008).

Another co-factor of the DNA methyltransferases is Lsh (sometimes called HELLS - helicase lymphoid-specific). Lsh is a putative member of the Snf2 helicase family, a protein family frequently involved in nucleosome remodelling (Geiman *et al.*, 1998; Dennis *et al.,* 2001). Lsh is expressed at high levels in lymphoid cells but can be detected at low levels in many tissues (Geiman *et al.,* 2001). Lsh first became associated with DNA methylation when it was observed that *Lsh*$^{-/-}$ mice are hypomethylated at certain loci (Dennis *et al.,* 2001; Sun *et al.*, 2004). Mice deficient in *Lsh* die perinatally or within weeks of birth, are runted compared to wildtype littermates and show premature aging phenotypes (Geiman *et al.,* 2001; Sun *et al.,* 2004). Hypomethylation in *Lsh*$^{-/-}$ mice was originally reported to occur at repetitive regions and some single copy loci (Dennis *et al.*, 2001). It was subsequently shown that repeats are a major target of Lsh as, in one study using expression microarrays, two thirds of upregulated transcripts in the absence of Lsh were suggested to originate from repetitive regions (Huang *et al.,* 2004). DNA methylation changes have been mapped in higher detail in the absence of Lsh which suggested that it is required to both promote and prevent DNA methylation of large chromosomal regions including many gene promoters (Myant *et al.*, 2010; Tao *et al.,* 2011a). As Lsh contains Snf2 and helicase domains, frequently involved in chromatin remodelling, an attractive function for Lsh in DNA methylation is to move nucleosomes to facilitate access by DNA methyltransferases. Indeed, nucleosome covered regions of DNA show lower levels of methylation than non-covered regions and are methylated at a lower efficiency by Dnmt3a *in vitro* (Felle *et al.*, 2011b). However, to date, no catalytic activity of Lsh on nucleosome templates has been reported. Instead, it has been suggested that Lsh is required for *de novo* methylation by interaction with Dnmt3a and Dnmt3b (Zhu *et al.*, 2006). The precise function of Lsh in DNA methylation remains unknown but it may involve targeting of Dnmt3s to genomic regions as, for example, Lsh associates with intact heterochromatin at pericentric regions (Yan *et al.*, 2003a). This model would be compatible with the large domains of DNA methylation that change in the absence of Lsh (Tao *et al.*, 2011a). Another possibility is that Lsh is required for maintaining or establishing a certain chromatin state that then secondarily influences DNA methylation levels. This idea is consistent with the observation that levels of histone modifications are altered in the absence of Lsh (Yan *et al.*, 2003b). Although it appears to be an important co-factor, further research is required to firmly establish the role of Lsh in DNA methylation.

G9a (otherwise known as Ehmt2 - euchromatic histone methyltransferase 2) has been implicated in DNA methylation. G9a is a histone methyltransferase responsible for a large amount of H3K9me2/3 (particularly K9me2) in euchromatic regions of the genome and is essential for embryonic development in mice (Tachibana *et al.,* 2002). Studies of the pluripotency-associated *Oct4 (Pou5f1)* gene upon embryonic stem cell differentiation showed that G9a is required for accumulation of H3K9me2/3 and DNA methylation at the promoter region during the silencing of this gene (Feldman *et al.,* 2006). One role of G9a at this locus appears to be in recruiting Dnmt3 proteins, as catalytically inactive G9a could partially rescue the failure to acquire full levels of DNA methylation in *G9a*$^{-/-}$ ES cells (Epsztejn-Litman *et al.*, 2008). The role for G9a in targeting DNA methylation appears to be more widespread, acting at various repetitive elements and gene promoters (Dong *et al.,* 2008; Tachibana *et al.*, 2008). It was reported that ~150 promoter regions are DNA hypomethylated in *G9a*$^{-/-}$ ES cells, many of which overlap with those hypomethylated in the absence of Lsh, suggesting a possible cross-talk between the two proteins (Myant *et al.*, 2010). Several other proteins have also been suggested to modulate DNA methylation patterns in mammals, such as the SWI/SNF containing ATRX protein (Gibbons *et al.,* 2000). An understanding of the full complement of DNA methylation co-factors is necessary to fully understand this epigenetic mark.

1.2.3. 5-Hydroxymethylation and other forms of modified cytosine

Recently, methyl-cytosine (5mC) was joined, as a major modified form of cytosine in mammalian genomes, by hydroxymethyl-cytosine (5hmC). Hydroxymethyl-cytosine consists of a hydroxymethyl group added to the 5' carbon of the pyrimidine ring of cytosine, the same position that is modified in methyl-cytosine (Fig. 1.1). The 5hmC base was rediscovered recently where it was found to be relatively abundant (~40% as abundant as 5mC as a neighbour of G) in Purkinje neurons of the brain (Kriaucionis and Heintz, 2009). The global abundance of 5hmC is variable between tissues and does not always correlate with global 5mC abundance (Li *et al.*, 2011c; Ruzov *et al.*, 2011; Nestor *et al.,* 2011). It was predicted that mammalian enzymes of the TET (ten eleven translocation) family could convert 5mC to 5hmC by oxidation and this was shown to be true for TET1 (Tahiliani *et al.*, 2009). Subsequently, a flurry of papers have emerged detailing the distribution and function of 5hmC and the TET proteins. A brief overview is given here followed by further discussion of 5hmC function and distribution in later sections.

There are three proteins of the TET family in mammals, Tet1, Tet2 and Tet3, and all have been implicated in conversion of 5mC to 5hmC in mammalian cells (Ito *et al.,* 2010). Tet1 was identified as being highly expressed in ES cells but its function here is unclear. Knockdown of *Tet1* in mouse ES cells reduced 5hmC, altered gene expression and influenced differentiation of ES cells, promoting trophectoderm differentiation, suggesting a role for the mark in pluripotency and differentiation (Ito *et al.,* 2010; Ficz *et al.,* 2011; Koh *et al.,* 2011; Williams *et al.,* 2011; Xu *et al.,* 2011b). However, *Tet1*$^{-/-}$ ES cells show only subtle changes in gene expression and *Tet1*$^{-/-}$ mice are viable displaying only a modest reduction in birth weight, suggesting that Tet1 is not required for pluripotency *in vivo* (Dawlaty *et al.,* 2011). A differentiation skew towards trophectoderm was observed for *Tet1*$^{-/-}$ ES cells *in vitro* in this report, consistent with the earlier knockdown studies. Only a modest reduction in 5hmC levels were observed in *Tet1*$^{-/-}$ ES cells suggesting that other enzymes contribute to the abundance of this modified base, and, that this modest reduction does not lead to major defects in embryogenesis (Dawlaty *et al.,* 2011). Tet2 is implicated in haematopoiesis and mutations in *TET2* are common in certain types of blood cancer (Ko *et al.,* 2010). Mutation of *Tet2* in mice results in differential maintenance and differentiation of hematopoietic stem cells (Ko *et al.,* 2011; Li *et al.,* 2011c; Quivoron *et al.,* 2011). Phenotypes out with of the hematopoietic system have not been reported to date in mice lacking *Tet2*, suggesting that it is not required for general embryonic development. Tet3 has been implicated in catalysing 5hmC of paternal DNA in the zygote but also appears to be expressed in many human somatic tissues suggesting a wider function (Wossidlo *et al.,* 2011; Gu *et al.,* 2011; Nestor *et al.,* 2011). Expression of *TET* genes and hmC patterns are variable across human tissues and it is currently unclear what the contribution of each TET protein is (Nestor *et al.,* 2011). Mutation or depletion of multiple Tet proteins simultaneously will further shed light on the importance of this reaction and the 5hmC base *in vivo*.

It is currently not clear whether 5hmC is functional in its own right or is a non-functional chemical intermediate of DNA metabolism as some lines of evidence link 5hmC to DNA demethylation (discussed in 1.2.4.9). Studies investigating the difference in binding and activity of proteins on 5mC- and 5hmC-containing DNA have begun to shed light on this question. Firstly, it has been shown that Dnmt1 maintenance methylation activity *in vitro* does not occur on hemi-methylated 5hmC templates (Valinluck and Sowers, 2007). A recent report provides evidence suggesting that 5hmC is lost in a replication-dependent manner in mouse pre-implantation embryos, supporting a model of 5hmC-mediated passive demethylation (Inoue and Zhang, 2011). However, the maintenance activity of Dnmt1 *in*

vivo is dependent on the co-factor Uhrf1, the SRA domain of which can bind both 5mCpG and 5hmCpG with equal efficiency *in vitro* (Frauer *et al.*, 2011). This means that speculation about 5hmC-mediated passive demethylation by prevention of maintenance methylation need to be suspended until further investigation is conducted. Secondly, the binding of some methyl-CpG-binding proteins, for example MeCP2, does not occur on 5hmC suggesting that this base may function to antagonise binding of certain methyl-CpG-binding proteins to DNA (Frauer *et al.*, 2011). To date, no protein has been identified that binds specifically to 5hmC but no doubt this will be a subject of intense research in the near future.

Recently, two further forms of modified cytosine have been detected in mouse tissues, 5-Formylcytosine (5fC) and 5-Carboxylcytosine (5caC)(Ito *et al.*, 2011; He *et al.*, 2011a). These bases were detected at very low levels (<20 bases for every million cytosines) and may be intermediates produced by Tet enzymes in a putative process of demethylation (Ito *et al.*, 2011). The putative role of Tet enzymes and 5hmC in the removal of 5mC as well as the genomic distribution of 5hmC will be discussed in later sections (1.2.4.8 and 1.2.4.9).

1.2.4. The distribution of DNA methylation

Understanding where the 5mC and 5hmC bases are located in the genome is key to understanding their function in development and disease. The development of large-scale DNA methylation mapping technologies, often coupled with next generation sequencing technologies, has greatly increased our knowledge of the distribution and dynamics of modified cytosine bases in multiple organisms (for example, Illingworth *et al.*, 2008). This section aims to summarise that knowledge in the context of understanding the function of DNA methylation. I will focus on DNA methylation patterns in mammals and will summarise insights into patterns in other organisms later in this section.

Following the observation that 5hmC can be relatively abundant in mammalian DNA, it was disturbing to learn that many of the techniques used to map 5mC cannot distinguish between this modified base and 5hmC. These techniques include most where bisulfite treatment is used to distinguish 5mC and unmodified cytosine, the use of many restriction enzymes and also certain protein domains used for enriching methylated DNA (Nestor *et al.*, 2010). This means that a caveat must be added to all the data generated using these techniques that formally we cannot distinguish between these two forms of modified cytosine.

A simplified representation of our current knowledge of the mammalian DNA methylome is shown in figure 1.4. The pattern and dynamics of methylation at some of these regions will now be discussed in more detail.

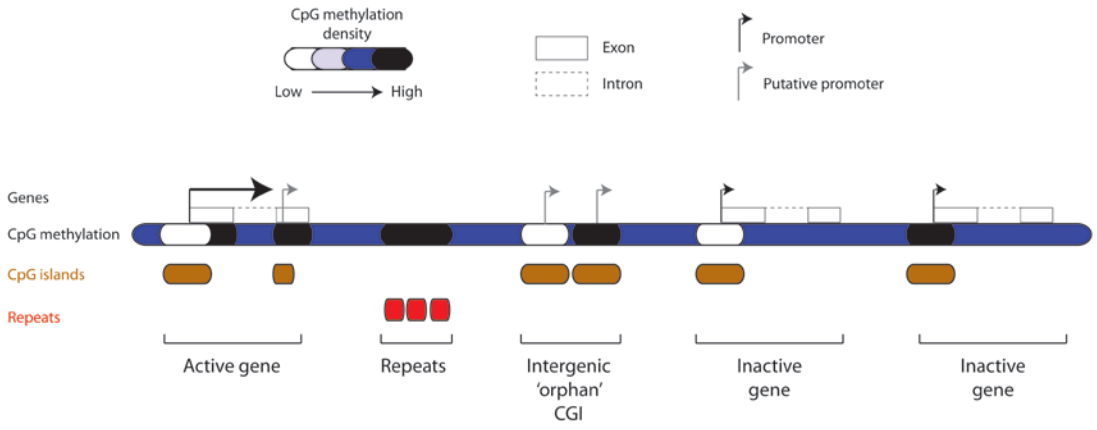


Figure 1.4 - The 5mCpG methylation landscape of mammalian genomes. 5mCpG methylation in vertebrate genomes is generally bimodal, with high methylation as the background level punctuated by low level DNA methylation at CpG islands. CpG islands are often associated with the 5' region of genes and these rarely have high levels of DNA methylation with the exception of those on the inactive X chromosome in female cells. A large proportion of CpG islands are not associated with annotated gene promoters (orphan CpG islands) and these are methylated more frequently and may represent putative promoters of unknown function (see 'intergenic orphan CGI') or alternative promoters of genes (see 'active gene'). Repeats and retrotransposons generally have high levels of DNA methylation and some contain an enrichment for CpG dinucleotides (as represented in the diagram). Many inactive genes are associated with a CpG island at their 5' end and these, like active genes, are usually unmethylated and are repressed through other mechanisms (inactive gene - left). A small proportion of promoter CpG islands have high levels of DNA methylation in a given cell type and this correlates with transcriptional repression (inactive gene - right). Gene bodies are often methylated, with exons showing higher DNA methylation than introns, and this may positively correlate with transcriptional activity (see 'active gene'). CGI: CpG island. The proportions of each feature in this diagram are of course not representative of the genome as a whole.

1.2.4.1. CpG islands

The vast majority of 5mC in mammals occurs at CpG dinucleotides (Lister *et al.*, 2009). Therefore, to understand the distribution of DNA methylation we first need to understand the distribution of its substrate within the genome. CpG dinucleotides are not distributed evenly throughout the mammalian genome. Instead they are found clustered in short stretches of DNA called CpG islands (Cooper *et al.*, 1983; Bird *et al.*, 1985). CpG islands are on

average ~1kb long and have a CpG-content and GC-content higher than the bulk genome. The commonly used sequence-based criteria for identifying a CpG island is at least 200bp of sequence with an observed/expected CpG content of greater than 0.6 and a GC content greater than 50%, although the stringency of this method can be adjusted by altering the criteria (Illingworth *et al.*, 2009). This method combined with repeat masking to discard repetitive sequences is used by the UCSC genome browser and yields about 28,000 islands in the human genome (Illingworth *et al.*, 2009). The CpG content of CpG islands (at 0.6 observed/expected) is not enriched compared to what is expected by chance but instead is merely less depleted than the bulk genome.

The pattern of CpG methylation in the mammalian genome is generally bimodal, as CpGs in the bulk of the genome are usually methylated and CpGs in CpG islands are usually unmethylated (Fig. 1.4)(Weber *et al.*, 2007; Meissner *et al.*, 2008). It was suggested that this pattern of methylation contributes to the depletion of the CpG dinucleotide that is observed in vertebrate genomes, as spontaneous deamination of methyl-cytosine to uracil, if not repaired by the mismatch repair machinery, will lead to its replacement by thymine (Bird *et al.*, 1985; Weber *et al.*, 2007). It is now suggested that this process, along with multiple other forms of selection, has lead to the distribution of CpG dinucleotides in living vertebrates (Cohen *et al.*, 2011). In contrast to purely sequence based classification, CpG islands have also been identified more functionally, by mapping DNA binding by an unmodified cytosine binding protein (Illingworth *et al.*, 2008; Illingworth *et al.*, 2010). For a CpG island to be identified in these studies it must have low levels of DNA methylation in at least one tissue studied, making the list of CpG islands identified non-comprehensive. These studies reconciled the number of CpG islands in mouse and human DNA, as sequence based methods had predicted far fewer CpG islands in the mouse genome. They showed that about half of CpG islands in mouse and human genomes are associated with the 5' end of annotated genes, and the other half were found in intra- or intergenic regions and were termed 'orphan' CpG islands. About 70% of protein coding genes were associated with a CpG island and 44% were associated with a CpG island at their 5' end. Unmethylated CpG islands show many properties that make them suited for a function at a gene promoter including nucleosome instability, depleted H1 occupancy and histone marks associated with a more open chromatin configuration (Tazi and Bird, 1990; Thomson *et al.*, 2010). This has lead to the idea that CpG islands facilitate transcription by differentiating gene promoters from bulk chromatin. Unmethylated CpG islands are also associated with the promoters of inactive genes, like tissue-specific developmental regulators, that need to be tightly regulated in time and space (Fig. 1.4). These genes are targets for repression by the polycomb repressor

complexes which have been suggested to use CpG islands for their recruitment (discussed in section 1.3). So, it appears that CpG islands are an integral part of a chromatin based switch for regulating gene expression.

Only a small proportion of CpG islands are highly methylated in normal tissues (around 10%)(Illingworth *et al.,* 2008; Deaton *et al.,* 2011). In particular, promoter-associated CpG islands are rarely methylated (<3% of them), but methylation, when present, does correlate with transcriptional silencing (Illingworth *et al.,* 2008; Maunakea *et al.,* 2010; Deaton *et al.,* 2011). Interestingly, orphan CpG islands are more often found to be methylated than promoter-associated CpG islands in normal tissues (intragenic CpG islands ~26% are methylated, intergenic ~ 13% in any of the tissues studied)(Illingworth *et al.,* 2008; Illingworth *et al.,* 2010; Maunakea *et al.,* 2010). The function of orphan CpG islands is unclear but many of them show evidence of promoter activity and therefore may be promoters for unannotated transcripts (Illingworth *et al.,* 2010; Maunakea *et al.,* 2010). If the methylation status of CpG islands is important for maintaining a transcriptional memory, then one would expect to find differential patterns of methylation at CpG islands in different tissues. In one study, only ~3% of promoter associated CpG islands showed differential methylation between blood, brain, muscle and spleen. This compared to 6.6-10.1% of orphan CpG islands that show differential methylation (Illingworth *et al.,* 2008). Although the proportions are small, the differentially methylated CpG islands were enriched for locations near developmentally important genes such as *Hox* and *Pax* genes. In another study looking at different cell lineages of the immune system, less than 0.5% of promoter associated CpG islands showed differential methylation compared to 0.6-2% of orphan islands (Deaton *et al.,* 2011). Again, many of the differentially methylated islands were located near genes of importance in the immune system. In this study a larger number of differentially methylated regions was found between brain and immune cells suggesting that more distantly related tissues may have more differential methylation. In addition, a larger number of differentially methylated regions were found outside of CpG islands. The methylation status of CpG islands may not always be uniform as blocks of hypermethylated and hypomethylated regions have been observed within the same CpG island (Hodges *et al.* 2009; Zhang *et al.,* 2009). Many of these blocks may correspond to underlying features such as exons but methylation heterogeneity within CpG islands will need to be explored in more detail.

The data from Illingworth *et al.,* (2008; 2010) and Deaton *et al.,* (2011) were generated using a CXXC domain from Mbd1 to bind unmethylated CpGs and a MBD (methyl-CpG binding domain) from the rat MeCP2 protein to bind methylated DNA (Nan *et*

al., 1993; Cross *et al.*, 1994; Illingworth *et al.*, 2008). It is currently unknown if this CXXC domain binds to 5hmC, although this seems unlikely based on structural studies of CXXC-DNA interactions, which are thought to be sterically inhibited by the 5mC modification (Cierpicki *et al.*, 2010). The MBD domain of mammalian MeCP2 has been shown to not bind, or bind with greatly reduced efficiency, to 5hmC compared with 5mC (Frauer *et al.*, 2011). This suggests that data collected using the MAP technique probably does not suffer from 5hmC contamination while for the CAP technique, it remains untested.

Together, these data suggest that the vast majority of CpG islands in the mammalian genome are not dynamically methylated and remain with low levels of DNA methylation in all tissues. Methylation of a small number of CpG islands does change in a tissue-specific manner and these changes may be important in differentiation. Interestingly, orphan CpG islands show higher variance in DNA methylation levels between tissues. The methylation dynamics at CpG islands, particularly orphan islands, will have to be studied in more detail in order to elucidate their function.

1.2.4.2. CpG island 'shores'

In an independent study, using a different experimental approach, tissue-specific methylation patterns were found to occur infrequently within CpG islands and gene promoters (Irizarry *et al.*, 2009). Instead, most variation in DNA methylation between tissues was observed at regions flanking CpG islands, termed 'CpG island shores'. These regions also showed the most variation, in studies by the same group, between normal colon and colon cancer DNA and also between human induced pluripotent cells and their parental fibroblasts (Irizarry *et al.*, 2009; Doi *et al.*, 2009). CpG island shores were defined as regions flanking a CpG island, within 2kb either side, and so have lower CpG content relative to their neighbouring CpG island (Irizarry *et al.*, 2009). The functional consequence of CpG shore methylation is unclear but methylation of CpG shores correlates with lower transcription of the nearby gene (Irizarry *et al.*, 2009; Ji *et al.*, 2010). Study of candidate genes showed that differential methylation at CpG shores correlates with the use of alternative promoters in some cases, suggesting that this process may influence the regulation of transcription and promoter choice (Irizarry *et al.*, 2009). Further support for a functional consequence of CpG shore methylation comes from the observation that human CpG shore methylation patterns are often conserved in equivalent mouse tissues (Irizarry *et al.*, 2009).

The same approach was applied to lineage committed and progenitor cells of the hematopoietic system in order to study the dynamics of DNA methylation during lineage

commitment (Ji *et al.*, 2010). Again, most differentially methylated regions were found in CpG shores and their methylation negatively correlated with the expression status of the nearby gene. In addition, the methylation status of shores correlated better with transcriptional activity than the methylation status of islands (Ji *et al.*, 2010). These correlations, however, do not take into account the large number of genes associated with an unmethylated CpG island that are silenced by other mechanisms, such as polycomb repressor complexes. In addition, the data on shores and transcription are so far only correlative. It must also be considered that the technique used to generate this data, digestion of 5mC with restriction enzyme McrBC, also cuts at 5hmC. This, together with the observation that 5hmC may be present at promoter and gene body regions (discussed in 1.2.4.8), means that a caveat must be added to this data currently that the observed variation may be at least in part due to variation in the 5hmC base. Nevertheless, further study of CpG island shores will illuminate their putative function in transcriptional regulation.

1.2.4.3. Gene promoters

As CpG methylation is linked to transcriptional regulation many studies have focussed on gene promoters when mapping DNA methylation. As a large proportion of genes in mammalian genomes are associated with a CpG island at their 5' end, much of the data discussed in the previous section on CpG islands is also relevant here. Similarly to CpG islands, the majority of promoter regions have low level methylation (for example, ~10% of promoters in B cells are highly methylated) and the methylation (5mCpG) on average at promoters is lower than surrounding regions (Rauch *et al.*, 2009; Borgel *et al.*, 2010). Importantly, the methylation level in promoter regions and in first exon has been anti-correlated with transcription in many studies (Eckhardt *et al.*, 2006; Shen *et al.*, 2007; Weber *et al.*, 2007; Meissner *et al.*, 2008; Rauch *et al.*, 2009; Borgel *et al.*, 2010; Brenet *et al.*, 2011). These data have lead to the generally accepted view that high levels of DNA methylation in a CpG-rich promoter region can attenuate its activity.

The density of CpGs in a gene promoter has been shown to predict both its methylation status and its relationship with transcriptional status (Weber *et al.*, 2007; Meissner *et al.*, 2008). Promoters can be divided into groups based on their CpG content. Weber *et al.* (2007) defined high-density CpG promoters (HCPs), intermediate-density CpG promoters (ICPs - sometimes known as 'weak CpG island promoters') and low-density CpG promoters (LCPs) based on sequence parameters. By mapping DNA methylation in human somatic and germline cells, they found that LCPs were nearly always highly methylated,

regardless of their transcriptional activity (Weber *et al.,* 2007). Conversely, HCPs were nearly always unmethylated, but the rare event of high methylation of a HCP correlated with transcriptional repression. These observations are in accordance with CpG island profiling that suggests that most CpG rich regions have very low levels of DNA methylation in all tissues. Also, these data are in accordance with the suggestion that a specific density of CpGs is required to have a significant silencing effect when methylated. ICPs were particularly interesting as they were more often found to be highly methylated and their promoter RNA polymerase II (polII) association was inversely correlated with DNA methylation status (Weber *et al.,* 2007). HCPs and ICPs were found to contain high levels of the histone mark H3K4me2 which is also inversely correlated with DNA methylation. One explanation of these data is that something, potentially a histone mark such as H3K4me2, protects regions with high CpG density from DNA methylation, and ICP promoters are more often subject to DNA methylation due to their reduced CpG content. ICPs have also been suggested to undergo hypermethylation more often than other promoter classes in cell culture suggesting that they are prone to *de novo* methylation (Meissner *et al.,* 2008). ICPs could therefore represent a balance between having enough CpGs so that, when methylated, they are capable of repressing transcription, and having not too high a CpG content that would be somehow refractory to methylation. There are of course additional determinants of DNA methylation, as, for example, many CpG rich promoters of germline genes are found to be highly methylated in somatic cells (Weber *et al.,* 2007). This data (Weber *et al.,* 2007) was generated using methyl-DNA immunoprecipitation (MeDIP) using an antibody that does not recognise 5hmC.

Profiling DNA methylation in combination with histone modifications at gene promoters has shown that the two types of chromatin mark are often highly correlated. In fact, it has been suggested that histone modifications are a better predictor of DNA methylation status than DNA sequence (Meissner *et al.,* 2008). For example, the presence of H3K4me3 at HCPs is negatively correlated with DNA methylation status (Meissner *et al.,* 2008; Hawkins *et al.,* 2010).

As with CpG islands, tissue specific patterns of DNA methylation at promoters have been investigated. Between 12 human tissues, 17% of promoter regions were found to be differentially methylated (Eckhardt *et al.,* 2006). Many promoter regions were shown to be differentially methylated between pluripotent ES cells and somatic fibroblasts suggesting dynamic methylation during differentiation (Meissner *et al.,* 2008). Promoter DNA methylation has been mapped during early mouse development suggesting that many promoters (476 genes) are *de novo* methylated between the blastocyst stage and E9.5 (Borgel

et al., 2010). A large proportion of these genes appear to be associated with ICP promoters and are expressed and have functions in the germline (Borgel *et al.*, 2010). This indicates that promoter methylation may be important for silencing germline gene expression programs in somatic cells (discussed in 1.2.5.5).

These studies indicate that promoter regions can be differentially methylated between tissues, a key property if they are to be involved in maintaining patterns of transcription.

1.2.4.4. Gene body methylation

Unbiased methods for DNA methylation mapping have allowed attention to move away from a promoter-centric approach which had previously dominated. One of the exciting new avenues of research that these studies have provided is gene body methylation. As vertebrate methylomes belong to the class of 'globally methylated', it may be no surprise that gene bodies are also methylated (Fig. 1.4). However, provocative evidence suggests that gene body methylation may represent a crucial function of DNA methylation that is relatively poorly understood. Gene body methylation is evolutionarily ancient and many organisms that have a so called 'mosaic' methylome exhibit gene body methylation (discussed in 1.2.4.10). For example, in the model flowering plant, *Arabidopsis thaliana*, the vast majority of DNA methylation (5mC) occurs at gene bodies and transposon sequences (Zilberman *et al.*, 2007). Furthermore, methylation is more prominent near the centre of the gene body than the ends (Zilberman *et al.*, 2007). In mammals, many tissue-specific differentially methylated regions have been found in gene bodies, including intragenic CpG islands, suggesting a function in gene regulation (Eckhardt *et al.*, 2006; Illingworth *et al.*, 2010; Maunakea *et al.*, 2010).

The function of gene body methylation is currently unclear. Many studies have found positive correlations between gene body methylation and transcription in mammals (Rauch *et al.*, 2009; Ball *et al.*, 2009; Laurent *et al.*, 2010; Aran *et al.*, 2011). In addition, an elegant study of the active (Xa) and inactive (Xi) X chromosomes in female cells showed that the Xa, where many genes are expressed, contains methylation concentrated at gene bodies, whereas the Xi was depleted of methylation at gene bodies (Hellman and Chess, 2007). In one study, this correlation was observed only in highly proliferative tissues linking it to actively dividing cells (Aran *et al.*, 2011). These data suggest that gene body methylation may be somehow linked to transcription. In contrast, the first single base resolution methylation map in mammals showed no correlation between gene body CpG

methylation and transcription in human ES cells, although a weak correlation was observed for fibroblast cells (Lister *et al.,* 2009; Hawkins *et al.,* 2010). In *Arabidopsis*, genes that are moderately transcribed are the most methylated while both high and low expressed genes contain lower methylation (Zilberman *et al.,* 2007). In addition, reduction of methylation in *Arabidopsis* leads to upregulation of genes with gene body methylation (Zilberman *et al.,* 2007). DNA methylation has also been suggested to inhibit transcriptional elongation in the filamentous fungi, *Neurospora crassa* (Rountree *et al.,* 1997). In mammals, certain histone marks associated with a more closed chromatin structure, such as H3K36me3 and H3K9me2/3, are associated with gene bodies of active genes (Barski *et al.,* 2007). In fact, H3K36me3 is considered a marker of actively elongating transcription units and a link between this modification and *de novo* DNA methylation has been established (discussed in 1.2.4.11). Therefore, it is feasible that DNA methylation could contribute to a more closed chromatin structure at gene bodies. DNA methylation may be required specifically in dividing cells where a memory of chromatin state across cell divisions must be maintained, but not required in quiescent cells where chromatin states may be more stable. But what is the function of a more closed chromatin conformation at gene bodies? One of the functions could be to prevent spurious transcription initiation. A speculation is that as the act of transcription ejects nucleosomes, leading to a more open chromatin conformation, this could promote new transcription initiation events producing cryptic transcripts. This could explain why in *Arabidopsis* moderately expressed genes contain highest gene body methylation, as at highly transcribed genes the next polymerase runs through the gene without time for reassembly of chromatin into a more closed state. Some evidence does exist for a role of DNA methylation in prevention of transcription initiation in gene bodies. A tissue-specifically methylated CpG island was shown to act as an alternative promoter at the *Shank3* gene, in both mouse and human, the activity of which is negatively regulated by DNA methylation (Maunakea *et al.,* 2010). However, a general role for gene body methylation in prevention of spurious transcription initiation within transcription units has not been tested and remains speculation.

1.2.4.5. Enhancers

Another finding of DNA methylation mapping studies is that enhancers may be differentially methylated, correlating with their activity. The binding sites for certain pluripotency-associated transcription factors in human ES cells were shown to be depleted of DNA methylation (Lister *et al.,* 2009). Some were found to be more depleted of methylation in ES

cells, where these factors are highly expressed, and less depleted in fibroblasts, where these factors are not expressed (Lister *et al.*, 2009). In addition, using a chromatin signature to differentiate active from inactive enhancers, it was shown that DNA methylation depletion was evident in fibroblasts at enhancers specifically active in these cells, although the same was not true for ES cell-specific enhancers (Lister *et al.*, 2009). This analysis was extended and many signatures of differential DNA methylation were found at cell-specific enhancers (Hawkins *et al.*, 2010). DNA methylation at enhancers has not been extensively studied and it is unclear from this data what the cause and effect relationships are between DNA methylation and enhancer activity. For example, it is feasible that DNA methylation is capable of blocking the binding of a DNA binding protein directly or contributing to a chromatin state that is incompatible with binding therefore precluding the activity of an enhancer (see later section 1.2.6 on the molecular function of DNA methylation). It is also feasible that enhancer binding proteins can prevent access to DNA by DNMTs and therefore lead to reduction in methylation at an active enhancer, or that a similar effect is achieved by a chromatin state that is refractory to DNA methylation. More study is needed in this respect to establish a potential role for DNA methylation in enhancer activity. Interestingly, a study on glucocorticoid receptor binding at enhancers showed that DNA methylation prevents binding and formation of a DNaseI hypersensitive site upon stimulation with glucocorticoids (Wiench *et al.*, 2011). Furthermore, knockdown of *Dnmt1* increased binding of the receptor to previously precluded sites suggesting that methylation may act to regulate tissue-specific activity of certain enhancer elements.

1.2.4.6. Methylation status over large domains

Unbiased maps of DNA methylation also allow the study of differential methylation of large domains. In the first base-resolution methylation map of mammals it was noticed that amongst the high background of methylation were large regions of lower methylation present in somatic cells but not ES cells (Lister *et al.*, 2009). These regions were large (mean length of 153kb), covered a substantial proportion of each chromosome (mean of 38.4% of autosomes), had a mean methylation less than 70% and were termed partially methylated domains (PMDs). Interestingly, genes within fibroblast PMDs (partial methylation in fibroblast cells) were less expressed in fibroblasts than ES cells and were associated with certain repressive histone marks (Hawkins *et al.*, 2010). In another study, PMDs were identified in neuronal cells and shown to be associated with tissue-specifically expressed genes that were repressed in the neuronal cells studied (Schroeder *et al.*, 2011).

Conversely, genes found in broad domains of high methylation in neuronal cells were enriched for neuronal functions. These studies, together with gene body methylation, demonstrate that the association of DNA methylation with transcription is context and location dependent. Gene regulation is more frequently being linked to higher order levels of regulation and it will be interesting to see what future studies tell us about the function of methylation of large genomic domains.

1.2.4.7. Non-CpG methylation

Methylation of cytosines in the non-CpG context is abundant in plant genomes (Feng *et al.*, 2010). It was noticed that mouse embryonic stem (ES) cells also have significant DNA methylation in the non-CpG context providing an alternative to the view that DNA methylation in mammals is restricted to CpG dinucleotides (Ramsahoye *et al.*, 2000). Subsequently, the abundance of non CpG methylation was estimated in a selection of animals and plants (Feng *et al.*, 2010). Non-CpG methylation was found to be much more abundant in plant species and is relatively rare in animals with about 0.3% of CHG and CHH sites methylated (where H is A, C or T) in mouse embryo DNA (Feng *et al.*, 2010). Non-CpG methylation has been mapped in detail in certain mammalian tissues (Lister *et al.*, 2009; Laurent *et al.*, 2010, Feng *et al.*, 2010; Li *et al.*, 2010). It was observed that ES cells are enriched for non-CpG methylation with nearly a quarter of cytosine methylation in human ES cells in the context of non-CpG. In contrast, somatic cells such as mouse and human fibroblasts and human blood mononuclear cells have much lower non-CpG methylation content, with 99.98% of methylation in human fibroblasts in the CpG content (Lister *et al.*, 2009; Laurent *et al.*, 2010; Li *et al.*, 2010). In another study looking at candidate regions, mainly at differentially methylated imprinted regions, it was found that oocytes contain roughly 35% non-CpG methylation compared to CpG methylation while these regions were nearly exclusively marked by CpG methylation in sperm, blastocyst and the E12.5 embryo (Tomizawa *et al.*, 2011). This further suggests that non-CpG methylation is linked to the tissue of origin. It is currently unclear if oocytes have a genome wide increase in non-CpG methylation or if it is locus-specific.

In human ES cells non-CpG methylation is reportedly depleted from gene promoters and 5' UTRs (un-translated region) and more abundant in gene bodies and 3' UTRs, with CHH methylation being enriched in exons relative to introns and 3' UTRs (Lister *et al.*, 2009). Both types of non-CpG methylation in gene bodies are positively correlated with gene expression with a small bias for marking the antisense strand. Like CpG methylation,

non-CpG methylation was depleted at protein binding sites and negatively correlates with gene expression when present at promoters (Lister *et al.,* 2009). In fact, non-CpG methylation broadly correlates with CpG methylation and is generally depleted from CpG islands and promoter regions. In addition, genes reportedly marked by gene body non-CpG methylation are enriched for functions in RNA processing, RNA splicing and RNA metabolic processes (Lister *et al.,* 2009).

The tissue-specificity and high abundance of non-CpG methylation in ES cells is intriguing and intuitively suggests function. However, until further studies are conducted it remains possible that it is merely non-functional noise related to high *de novo* methyltransferase activity. *De novo* DNA methyltransferases are expressed highly in ES cells (Lister *et al.,* 2009; Laurent *et al.,* 2010). ES cells are derived from pluripotent tissues of the early embryo and are artificially maintained in an actively dividing and pluripotent state in cell culture that exists only transiently during development. In fact, prolonging the high levels of Dnmt3 proteins by culturing mouse ES cells is thought to lead to an altered epigenome relative to their tissue of origin (Borgel *et al.,* 2010). It is therefore entirely possible that extensive non-CpG methylation in hES cells, reported by Lister *et al.,* (2009), exists as a consequence of prolonged high levels of *de novo* methylation and serves no function. Convincing evidence to the contrary would be extensive bisulfite sequencing from early embryo DNA showing non-CpG methylation, demonstrating that it is not a consequence of cell culture. Very low levels of non-CpG methylation have been detected in somatic cell lines and tissues studied so far (Lister *et al.,* 2009; Laurent *et al.,* 2010; Feng *et al.,* 2010; Tomizawa *et al.,* 2011). Non-CpG methylation in hES cells is found enriched precisely at regions that are thought to recruit Dnmt3s, including actively transcribing gene bodies which may recruit Dnmt3s through the H3K36me3 mark (discussed in 1.2.4.11). As much as 98% of non-CpG methylation was found to be asymmetrical in hES cells, being present on only one strand, compared to 1% of CpG methylation that is asymmetrical (Lister *et al.,* 2009). This means that non-CpG methylation would not be maintained across cell divisions. For these reasons, it remains possible that ES cells are generally hypermethylated and non-CpG methylation is part of the noise associated with prolonged high Dnmt3 activity. Oocytes represent a non-cultured cell type that has reported non-CpG methylation, although its location and abundance on a large scale has not yet been reported. It is possible that as oocytes are non-dividing that non-CpG methylation accumulates as it is not diluted by DNA replication. It will be interesting to see the results of future studies of non-CpG methylation and whether, as in plants, it has a functional consequence.

1.2.4.8. Distribution of 5hmC

The development of multiple mapping techniques has allowed the distribution of 5hmC to be studied recently. Here I will summarise the findings of these initial studies, which have nearly exclusively been conducted using mouse ES cells. Both 5hmC and the binding of the Tet1 protein have been mapped in mouse ES cells. The 5hmC base has been found throughout the ES cell genome but is particularly enriched in gene bodies and promoter regions (Ficz *et al.,* 2010; Pastor *et al.,* 2011; Szulwach *et al.,* 2011; Williams *et al.,* 2011; Wu *et al.,* 2011; Xu *et al.,* 2011b). Genome-wide, 5hmC is fairly well correlated with 5mC with the exception of promoter and CpG-rich regions, which are depleted for 5mC but often enriched for 5hmC (Szulwach *et al.,* 2011). Like 5mC, the 5hmC signal at gene promoters is inversely correlated with transcription while within gene bodies it is positively correlated with transcription (Pastor *et al.,* 2011; Szulwach *et al.,* 2011; Wu *et al.,* 2011). However, the pattern of 5hmC at promoters may be more complex. At highly expressed genes, although 5hmC is not present over a punctate region at the transcription start site (TSS), it is present either side of the TSS, suggesting that it may be actively depleted from the TSS at these genes (Szulwach *et al.,* 2011). The Tet1 protein contains a CXXC domain and is strongly enriched for binding at CpG-rich regions of the genome (Williams *et al.,* 2011; Xu *et al.,* 2011b). 5hmC also shows a weak correlation with genomic CpG density although the locations of Tet1 and 5hmC are not always overlapping (Szulwach *et al.,* 2011; Williams *et al.,* 2011). For example, Tet1 is bound to CpG-rich promoters, independent of their activity, while 5hmC is present at the genes at the lower expression level at this region (Wu *et al.,* 2011). Further study of the location and turnover of the 5hmC base may resolve this discrepancy.

Most of the studies that have mapped 5hmC have used immunoprecipitation based techniques that enrich for DNA containing 5hmC. Therefore, not much is known about the absolute quantity of the 5hmC base at specific loci in mammalian genomes. One study suggests that 5hmC can represent as much as ~70% of all forms of cytosine at a single loci making it the most abundant form of cytosine at this site (Nestor *et al.,* 2011). Also, this study suggests that the absolute quantity of 5hmC at loci can vary greatly between tissues and does not correlate with either gene transcriptional status or expression of the *TET* genes, a feature that can be masked when using immunoprecipitation techniques (Nestor *et al.,* 2011).

As with any chromatin modification it is important to understand 5hmC in the context of the epigenome. Correlations have been established between 5hmC and certain

histone modifications in ES cells (Szulwach *et al.*, 2011). Genome wide, 5hmC correlates well with 5mC, H3K4me1 and H3K18ac while poor correlations were observed for other marks tested (Szulwach *et al.*, 2011). Tet1 and 5hmC can be found at many genes marked by binding of polycomb group proteins and H3K27me3 (Wu *et al.*, 2011; Pastor *et al.*, 2011). Particularly, 5hmC is enriched at promoters marked by both H3K4me3 and H3K27me3, also known as the bivalent mark (Pastor *et al.*, 2011). The significance of this association is unclear but may reflect a need to maintain low 5mC at these regions (Wu *et al.*, 2011). A peak of 5hmC has also been detected at various enhancers in ES cells (Szulwach *et al.*, 2011). More of this type of study will be important to determine the dynamics and putative function of the 5hmC base.

1.2.4.9. Developmental dynamics of DNA methylation and demethylation

DNA methylation patterns are remarkably dynamic during the mammalian life cycle but the mechanisms behind these dynamics have been enigmatic. Extensive demethylation is thought to occur during at least two stages of the mouse life cycle, in the zygote and in the developing germline. The DNA methylation patterns in mature sperm and oocytes are markedly different, with sperm having a generally higher 5mC content (Monk *et al.*, 1987; Kafri *et al.*, 1992; Smallwood *et al.*, 2011). Upon fertilisation, the paternal genome undergoes extensive loss of 5mC without replication of DNA suggesting that active demethylation may take place (Wossidlo *et al.*, 2011; Gu *et al.*, 2011). DNA methylation dynamics at this stage are not a general feature of vertebrates, with a variable extent of demethylation observed in different organisms, and no observable dynamics in others (Macleod *et al.*, 1999). In mammals, the zygotic wave of demethylation of the paternal genome has been suggested to be involved in restoring totipotency to the genome, and is implicated in removing methylation from the promoters of key pluripotency genes (Farthing *et al.*, 2008; Gu *et al.*, 2011). This demethylation is followed by extensive *de novo* methylation of the genome before the embryo implants, and also following implantation, leading to the global methylome that was described in a previous section (Borgel *et al.*, 2010).

The second stage of DNA demethylation in mice occurs during germ cell development. Germ cell precursors, primordial germ cells (PGCs), undergo DNA demethylation as part of genome-wide chromatin reprogramming phases during their migration to, and arrival in, the developing gonads (Hajkova *et al.*, 2008; Hajkova *et al.*, 2010). One of the functions of these epigenetic reprogramming events in germ cells is

thought to be removal of genomic imprints, parent of origin-specific DNA methylation marks, so that they can be re-established in a sex-dependent manner (Hajkova *et al.,* 2008). Another suggested function is in the establishment of a chromatin pattern that is compatible with future totipotency upon fertilisation although the full extent and precise function of these demethylation phases remain unclear (Hajkova *et al.,* 2008).

Active DNA demethylation has long been presumed to exist in mammals, although the mechanisms involved have remained largely enigmatic due to the absence of a clear DNA demethylase. Recent evidence suggests that the 5hmC base and DNA repair mechanisms may be at least part of the answer to this long standing question. Multiple DNA repair components have now been implicated in DNA demethylation, including Gadd45a (Barreto *et al.,* 2007), Mbd4 (Rai *et al.,* 2008; Kim *et al.,* 2009a) and TDG (Cortellino *et al.,* 2011; He *et al.,* 2011a). A deaminase enzyme, AID, has been implicated in demethylation of DNA, including during the reprogramming phase of PGC development (Popp *et al.,* 2010; Bhutani *et al.,* 2010). These data together suggested a model for demethylation involving deamination of 5mC to uracil followed by subsequent base excision mediated by a glycosylase such as Mbd4 or TDG. Evidence has been put forward for a role of Mbd4 in demethylation upon hormone-induced transcription, although Mbd4 is unlikely to play a major role in demethylation generally as *Mbd4*$^{-/-}$ mice are viable and fertile (Kim *et al.,* 2009a). TDG has emerged as a better candidate, as *Tdg*$^{-/-}$ mice show embryonic lethality and have increased DNA methylation at candidate loci (Cortellino *et al.,* 2011; Cortazar *et al.,* 2011). The re-emergence of the 5hmC base has lead to further insight into repair-mediated active demethylation. The 5hmC base accumulates on the paternal genome in the zygote during loss of 5mC, and the Tet3 protein is required for the observed dynamics of 5hmC and 5mC during this period (Wossidlo *et al.,* 2011). In addition, 5hmC and Tet proteins have been implicated in demethylation in somatic tissues such as brain (Guo *et al.,* 2011). TDG may be the link between 5hmC and demethylation. TDG has been implicated in excision of 5hmC following deamination by AID and also in excision of 5caC, a further modification of cytosine catalysed by Tet enzymes (Cortellino *et al.,* 2011; He *et al.,* 2011a). These data suggest models for the mechanism by which conversion of 5mC to 5hmC by Tet proteins initiates active demethylation. The 5hmC base may be deaminated by AID to 5'-hydroxymethyluracil then excised by TDG. Or, a deamination independent reaction may occur where Tet enzymes catalyse 5caC which is excised by TDG. These models are beginning to shed light on how DNA methylation dynamics are governed and also that active DNA demethylation may be a multi-step process involving Tet enzymes and DNA repair. However, a recent study suggests that the 5hmC that accumulates on the paternal genome in

the zygote is lost passively over the first few cell divisions of the early embryo and is not actively removed (Inoue and Zhang, 2011). Thus, although the dynamics of modified cytosine bases are becoming better understood, we are a long way from a good comprehension of the processes involved.

1.2.4.10. DNA methylation patterns in other organisms

Advances in methylation mapping technologies have led to a leap forward in our knowledge of DNA methylation patterns in other organisms allowing us to better understand the complex evolution of DNA methylation (Zemach *et al.,* 2010; Feng *et al.,* 2010). DNA methylation is conserved in members of most major groups of eukaryotes that have been studied including those of plants, fungi and animals. Some groups appear to have lost DNA methylation completely including the nematode *Caenorhabditis elegans* and the budding yeast *Saccharomyces cerevisiae*. Other clades appear to share some features of the DNA methylome and not others suggesting certain parts or functions were lost during their evolution. DNA methylomes can be classified as either global or mosaic depending on their pattern. An example of a global methylome is that of vertebrates where 5mC is the most prevalent form of cytosine in the CpG context and most CpGs in the genome are highly methylated (for example, Lister *et al.,* 2009). As previously mentioned, an example of a mosaic methylome is the flowering plant *Arabidopsis thaliana*, where most 5mC is present in domains at gene bodies and transposon sequences with the majority of the genome being low/unmethylated (Lister *et al.,* 2008). Two features of DNA methylomes appear to be evolutionarily ancient, gene body methylation and repeat/transposon methylation (Zemach *et al.,* 2010; Feng *et al.,* 2010). These two features were likely present in the common ancestor of eukaryotes and are present or absent in various combinations in extant eukaryotes (Zemach *et al.,* 2010). If DNA methyltransferases are grouped into three families based on homology (Dnmt1/MET1, Dnmt3/DRM and CMT/DIM-2 families), presence or absence of different groups is also somewhat related to these methylome features. For example, fungi, that have DNA methylation, have it at transposons but not gene bodies and have lost Dnmt3/DRM family members (Sukuki and Bird, 2008; Zemach and Zilberman, 2010). Invertebrate animals, that have DNA methylation, have it at gene bodies and not transposons and have lost the CMT/DIM-2 family member. The exception within animals is of course vertebrates which appear to have reacquired transposon methylation by switching to a global methylome (Lister *et al.,* 2009). Among plants there are species that have both gene body and transposon methylation and species that have only transposon methylation. This

suggests the following model for evolution of DNA methylation. The last common ancestor of eukaryotes likely had both gene body methylation and transposon methylation and had a representative DNMT from each of the three protein families (Zemach *et al.*, 2010). Subsequently, in different lineages, either or both features were lost and in some species DNA methylation was lost completely. Animals lost transposon methylation and seem to have adapted new methods for transposon silencing, such as repressive histone modifications, as seen in extant invertebrates (Zemach and Zilberman, 2010). Vertebrates seem to have reacquired transposon methylation by methylating nearly all of the genome. The reason for this switch to a global methylome in vertebrates is unclear but may have involved an increasing number of transposons in the population brought about by sexual reproduction. Indeed, the complex evolution of DNA methylation patterns remains a mystery but it has been suggested that sexual reproduction may have been a driving force (Zemach and Zilberman, 2010). DNA methylation seems to have been co-opted for gene regulation at promoter regions in plant and vertebrate species with promoter methylation negatively correlating with transcription in multiple species (Zemach *et al.*, 2010).

1.2.4.11. Targeting of DNA methylation

How specific DNA sequences are targeted for methylation remains unclear but current evidence supports a multifactorial system, including roles for transcription factors, chromatin proteins, histone modifications and non-coding RNA (ncRNA). Here I will give examples of each type of mechanism for targeting of DNA methylation. Given that vertebrates have a global methylome, with most of the CpGs in the genome highly methylated, with the exception of CpG islands, it is possible that they have evolved a 'methylation by default' targeting system. Such a system would entail, not methylation targeting per se, but protection of regions that need to remain unmethylated, such as the majority of CpG islands. Given recent data, it is interesting to speculate that a combination of H3K4me3 and the activity of Tet proteins contributes to the maintenance of a methylation-free state at CpG islands amongst the methylated bulk genome, although further work is needed to confirm these ideas. Also, DNA methylation has been suggested to be preferentially targeted to genomic regions that are transcriptionally silent, a process that may involve transcription- and repression-associated histone marks (discussed below)(Bird, 2002). Although it is possible that a system like this may operate in vertebrates, particularly during the *de novo* methylation periods of early embryogenesis, it is clear that there are also conserved mechanisms for targeting methylation to specific regions.

Circumstantial evidence links certain transcription factors to DNA methylation targeting. The pluripotency-associated *Pou5f1* gene (encoding the Oct4 protein) has been studied as a model for gene silencing during differentiation of embryonic stem cells. The promoter of *Pou5f1* is *de novo* methylated during this process, following transcriptional silencing (Feldman *et al.,* 2006; Epsztejn-Litman *et al.,* 2008). A nuclear receptor, GCNF (germ cell nuclear factor), is required for silencing and DNA methylation of the *Pou5f1* promoter and is reported to interact with Dnmt3 proteins (Sato *et al.,* 2006). Although this implicates GCNF in targeting of DNA methylation to this locus, it is possible that GCNF is required for the initial transcriptional silencing of the gene which in turn is required for DNA methylation. Further study would be required to elucidate the precise role of GCNF in *Pou5f1* silencing. Another example is E2F6, a transcriptional repressor implicated in targeting *de novo* methylation to specific germline-expressed genes in somatic cell lineages (Velasco *et al.,* 2010). Here, the Dnmt3b protein was shown to interact with E2F6 and *E2F6*$^{-/-}$ mice showed lack of Dnmt3b recruitment to, and reduced methylation of, target gene promoters (Velasco *et al.,* 2010).

DNA methylation patterns have been suggested to be better associated with histone modifications than the underlying genetic sequence (Meissner *et al.,* 2008). It is emerging that a crosstalk between histone modifications and the DNA methylation machinery is crucial in targeting DNA methylation. This crosstalk can occur at two levels, through protein-protein interactions between chromatin modifiers and DNMTs, or via a histone mark itself. Many chromatin associated proteins have been implicated in targeting of *de novo* methylation through protein-protein interaction, including G9a and Lsh that were discussed previously. The polycomb repressor complex 2 (PRC2) component, EZH2, has also been suggested to target DNA methylation in cancer cells by recruiting DNMTs to target genes (Vire *et al.,* 2006). However, this idea remains highly controversial as a similar effect has not been observed in normal tissues and is also not evident in all cancer cell lines tested (Denis *et al.,* 2011). Dnmt3 proteins appear to receive cues from histone modifications that influence their targeting. One such example is targeting by the Dnmt3L protein, that binds specifically to H3 tails unmodified at K4 via its PHD-like domain (Ooi *et al.,* 2007). The PHD-like domains of Dnmt3 proteins have also been suggested to be recognition modules for specific histone signatures. Similar to the PHD-like domain of Dnmt3L, the equivalent domains in Dnmt3a and Dnmt3b proteins bind preferentially to histone H3 tails that do not have methylation at the K4 position (Otani *et al.,* 2009; Zhang *et al.,* 2010b). The combination of Dnmt3L and Dnmt3s inhibition by H3K4me3 could explain part of the reciprocal locations of this histone mark and DNA methylation on chromatin (Meissner *et*

al., 2008). The PHD-like domain of Dnmt3a has also been implicated in recognising the H4R3me2 (symmetrically dimethylated arginine) modification catalysed by PRMT5 (protein arginine methyltransferase 5) and this interaction was suggested to influence DNA methylation at the silent β-globin locus (Zhao *et al.,* 2009). The PWWP domain of Dnmt3a has been shown to bind to H3K36me3 *in vitro* giving a potential explanation for the positive correlation between gene body methylation and transcription, and the binding of Dnmt3a to transcribed regions (Dhayalan *et al.,* 2010). H3K9me3 is correlated with DNA methylation at pericentric heterochromatin and has been implicated in targeting *de novo* methylation in three ways (Lehnertz *et al.,* 2003). Firstly, HP1 (heterochromatin protein 1), a K9me2/me3 binding protein, interacts with Dnmt3b (Lehnertz *et al.,* 2003; Geiman *et al.,* 2004). Secondly, Dnmt3b interacts with a H3K9 methyltransferase responsible for K9me at pericentric heterochromatin, Suv39h1 (Geiman *et al.,* 2004). Thirdly, DNA methylation may be targeted to H3K9me3 by Uhrf1. In addition to its role in maintenance of DNA methylation, Uhrf1 has been implicated in binding specifically to the combination of histone marks H3K9me3 and H3K4me0, and therefore may also be involved in directing DNA methylation to K9me3 (Rottach *et al.,* 2010; Nady *et al.,* 2011). These kind of mechanisms implicate the DNA methylation machinery as both 'readers' and 'writers' of chromatin modifications and underline the importance of chromatin crosstalk in gene regulation.

One emerging theme in DNA methylation targeting in mammals is RNA-mediated targeting. RNA-directed DNA methylation (RdDM) is better understood in plants where it is thought to account for a significant proportion of DNA methylation targeting (Lister *et al.,* 2008). In mammals, evidence for a similar process is starting to emerge. One class of small RNA, termed piRNA (piwi interacting RNA), has been implicated in targeting DNA methylation to transposons in the male embryonic germline (Aravin *et al.,* 2008; Kuramochi-Miyagawa *et al.,* 2008). piRNAs are small RNAs of 26-31 nucleotides in length that are classified by interaction with homologs of *Drosophila* Piwi proteins, proteins of the Argonaute family (Aravin *et al.,* 2008). piRNAs are thought to originate from transposon sequences, and after processing and amplification, feedback to target silencing mechanisms to repeat regions (Aravin *et al.,* 2008). Knockout of mouse homologs of Piwi proteins results in phenotypes similar to *Dnmt3l* mutation, including failure to silence transposon elements in the developing germline (Kuramochi-Miyagawa *et al.,* 2008). DNA methylation is reduced at transposon sequences in Piwi-homolog mutants in mice and specifically a role in *de novo* methylation during foetal germ cell development is suggested (Aravin *et al.,* 2008; Kuramochi-Miyagawa *et al.,* 2008). The molecular mechanism of DNA methylation targeting by piRNAs and Piwi proteins has not been established, and to date no interaction

between Piwi proteins and Dnmt3 proteins has been reported. As a piRNA-mediated transposon repression system exists in *Drosophila*, an organism with very low levels of DNA methylation, a sensible hypothesis might be that piRNAs guide other chromatin modifying complexes to transposons and DNA methylation is secondary to this event in mammals. Recently, RdDM in mammals was extended beyond piRNAs in a study of ribosomal DNA (rDNA) silencing (Schmitz *et al.,* 2010). A ncRNA complementary to the promoter of a rDNA gene, termed pRNA, was shown to form an RNA:DNA triplex structure capable of recruiting DNMT3B *in vitro*. Depletion or over-expression of pRNA led to changes in DNA methylation at the rDNA gene promoter implicating pRNA in targeting of *de novo* methylation (Schmitz *et al.,* 2010). As RNA could provide the crucial specificity, these mechanisms provide an intriguing system for DNMT targeting, and no doubt further studies will investigate a putative wider role for RNA:DNA triplexes and RNA based systems in general in targeting of epigenetic marks.

1.2.5. The roles of DNA methylation in mammals

The fundamental role of DNA methylation in mammals is generally thought to be in the repression of transcription. Due to its relative stability through DNA replication, it is implicated as a template for maintaining patterns of transcription through time and cell divisions, and is thought to be a major component of epigenetic memory. In this section I will discuss our knowledge of the roles of DNA methylation in transcriptional regulation during embryonic development and cell differentiation. I will then focus on the molecular mechanisms by which this epigenetic mark is thought to influence transcription.

1.2.5.1. DNA methylation is required for embryonic development

DNA methylation is required for mammalian development. This view has arisen as mice with mutations in genes required for normal DNA methylation show developmental phenotypes (detailed in table 1.1). In addition, a mouse with a catalytically inactive Dnmt1 shows roughly the same developmental phenotype as mice null for *Dnmt1*, suggesting that the DNA methylation mark itself, or at least the catalytic activity of Dnmt1, is required for development (Takebayashi *et al.,* 2007). Depletion or mutation of DNA methyltransferases in other organisms also leads to varying developmental defects and lethality, including

Xenopus laevis, zebrafish and *Arabidopsis thaliana* (Stancheva and Meehan, 2000; Stancheva *et al.*, 2002; Rai *et al.*, 2010; Mathieu *et al.*, 2007).

Protein	Major Function	Mutant phenotype (mouse)	Reference(s)
DNA methyltransferases			
Dnmt1	Maintenance methylation	N-allele (hypomorph): Mid-gestation lethality (10.5-12.5dpc); 10-20 somites; developmentally retarded (by 1-2 days); organ rudiments formed but small	Li *et al.*, 1992
		S- and C-alleles: Early to mid-gestation lethality (9.5-11.5dpc, 0-8 somites); developmentally retarded (by 1-2 days); distorted neural tube; failure to turn	Lei *et al.*, 1996
Dnmt3a	*De novo* methylation	Post-natal lethality (~4 weeks); runted	Okano *et al.*, 1999
Dnmt3b	*De novo* methylation	Embryonic lethality (14-18dpc); growth defects; neural tube defects	Okano *et al.*, 1999
Dnmt3a; Dnmt3b (double mutant)	*De novo* methylation	Early to mid-gestation lethality (9.5-11.5dpc); lack somites; developmentally retarded (by 1-2 days); distorted neural tube; failure to turn	Okano *et al.*, 1999
Dnmt2	tRNA methylation	None	Goll *et al.*, 2006
DNA methylation co-factors			
Dnmt3L	*De novo* methylation	Viable; male sterility	Bourc'his *et al.*, 2001
Lsh (Hells)	*De novo* methylation(?)	Peri-natal lethality (0- a few weeks); runted; premature aging phenotypes	Geiman *et al.*, 2001 Sun *et al.*, 2004
Uhrf1	Maintenance methylation	Early to mid-gestation lethality; similar to *Dnmt1*$^{-/-}$	Sharif *et al.*, 2007

Table 1.1 - Embryonic phenotypes of mice with mutated DNA methyltransferases and associated proteins.

The exact reason for failure of embryonic development in severely hypomethylated mice is unclear. Loss of genomic imprinting could have been a feasible explanation, as imprinted genes often have important developmental functions, and lack of imprints in oocytes, caused by *Dnmt3a* mutation, leads to offspring that die during mid-gestation (Kaneda *et al.*, 2004). However, mice lacking both *Dnmt3a* and *Dnmt3b* can reportedly maintain methylation at imprints during early embryogenesis and show the same phenotypes as *Dnmt1* loss (Okano

et al., 1999). This suggests that biallelic expression of imprinted genes is not the major cause of the developmental phenotypes, and also that *de novo* methylation is important for embryogenesis. A major contributor to the phenotype of *Dnmt1* loss is thought to be apoptosis, which is apparent in a significant proportion of cells in a *Dnmt1*$^{-/-}$ embryo and also in mouse embryonic fibroblasts lacking *Dnmt1* (Jackson-Grusby *et al.,* 2001; Lande-Diner *et al.,* 2007). Depletion of Dnmt1 in *Xenopus laevis* embryos also leads to induction of apoptosis, suggesting conservation of this response (Stancheva *et al.,* 2001). In addition to apoptosis, it seems plausible that transposon upregulation and gene mis-regulation contribute to the phenotypes.

1.2.5.2. DNA methylation is involved in cell differentiation and maintenance of cell identity

In addition to studies of embryonic development, cell lines and adult stem cell populations have been used to examine to importance of DNA methylation. Embryonic stem (ES) cells deficient in *Dnmt1* can be maintained in the undifferentiated state (Lei *et al.,* 1996). In fact, deletion of all three catalytically active DNA methyltransferases (TKO - for triple knockout) in ES cells completely abrogates DNA methylation but does not induce apoptosis or differentiation, indicating that DNA methylation is not required for ES cell self renewal (Tsumura *et al.,* 2006; Fouse *et al.,* 2007). However, normal levels of DNA methylation are required for another property of ES cells, pluripotency. ES cells deficient in *Dnmt1,* or both *Dnmt3a* and *Dnmt3b,* have defects in differentiation (Jackson *et al.,* 2004). *Dnmt1*$^{-/-}$ ES cells can initiate differentiation, measured by silencing of *Oct4* and loss of alkaline phosphatase staining, but show increased apoptosis and cannot terminally differentiate (Lei *et al.,* 1996; Jackson *et al.,* 2004). The same is observed for ES cells expressing a catalytically inactive Dnmt1 indicating that the DNA methylation mark itself is important during differentiation (Damelin and Bestor, 2007).

A number of studies implicate DNA methylation in the memory of lineage commitment after differentiation. An interesting example is the *Elf5* gene promoter, which is methylated in embryonic but not extra-embryonic tissues following the first cell lineage decision of the early embryo (Ng *et al.,* 2008). Methylation of this region is required for maintaining silencing of *Elf5* in the embryonic compartment, while in the extra-embryonic compartment *Elf5* initiates a positive feedback loop that reinforces extra-embryonic development. Therefore, ES cells lacking DNA methylation can contribute to the extra-

embryonic lineages when injected into a blastocyst while wildtype ES cells rarely achieve this (Ng *et al.*, 2008). DNA methylation has also been recently linked to the maintenance and differentiation of multipotent adult stem cell populations. *Dnmt1* is required to maintain multipotent progenitors of the epidermis and hematopoietic system in the undifferentiated state (Broske *et al.*, 2009; Trowbridge *et al.*, 2009; Sen *et al.*, 2010). In hematopoietic stem cells (HSCs), loss or depletion of *Dnmt1* results in cell differentiation and a skew towards myeloid, rather than lymphoid, cell lineages. Interestingly, this effect is independent of apoptosis and is thought to at least partially involve loss of repression of myeloid-specific genes (Broske *et al.*, 2009; Trowbridge *et al.*, 2009). Another study showed that promoter DNA methylation of *Arx*, a lineage-specific gene, in pancreatic β cells was necessary to maintain their identity. Deletion of *Dnmt1* resulted in loss of *Arx* repression and conversion to an α cell identity (Dhawan *et al.*, 2011). These studies provide evidence for the long standing hypothesis that DNA methylation, as a relatively stable epigenetic mark, is an important contributor to cell identity by influencing gene transcription.

1.2.5.3. X-inactivation and genomic imprinting: paradigms for transcriptional regulation by DNA methylation

Two biological processes have emerged as paradigms for transcriptional regulation by DNA methylation, X-chromosome inactivation and genomic imprinting. DNA methylation plays key roles in these mechanisms and they particularly exemplify its function in maintaining patterns of a transcriptional state. They are excellent models for studying the role of epigenetic mechanisms in transcriptional regulation for two reasons. Firstly, they involve allele-specific regulation, that cannot be explained by the presence or absence of trans acting transcription factors, as they exist within the same nucleus and thus are exposed to the same nuclear milieu. This means that the differential expression is likely governed by epigenetic factors. Secondly, they involve the memory of differential transcriptional states over long periods of time, in some cases for the life of the organism.

X-chromosome inactivation (XCI) is the process that results in large portions of one of the two X-chromosomes in female cells being transcriptionally inactivated (reviewed in Augui *et al.*, 2011). In the early female mouse embryo, both X-chromosomes are active until the 4-cell stage when the paternal X chromosome is inactivated (forming an inactive X - Xi) and this chromosome will remain inactive in the future extra-embryonic tissues. In the pluripotent cells of the blastocyst, that will give rise to the embryo, the paternal Xi is

reactivated followed by a process of random XCI that occurs around the epiblast stage (Augui *et al.,* 2011). The randomly chosen Xi is subsequently maintained in this state throughout all subsequent daughter cells for the life of the organism. Crucially, XCI involves heterochromatinisation of the Xi followed by changes in DNA methylation (for example, Heard *et al.,* 2001; Hellman and Chess, 2007; Sharp *et al.,* 2011). Relative to the Xa (active X chromosome), the Xi has lower levels of total DNA methylation, including at gene bodies, but many CpG islands associated with transcriptional start sites become highly methylated, mainly at the genes that are silenced by XCI (Hellman *et al.,* 2007; Sharp *et al.,* 2011; Suzuki and Bird, 2008). It is thought that DNA methylation is not required for the initial establishment of XCI but is important for its maintenance. Embryos that lack both *de novo* DNA methyltransferases can establish XCI but genetic or chemical perturbation of Dnmt1 in somatic cells results in reactivation of genes on the Xi (Sado *et al.,* 2000; Sado *et al.,* 2004; Csankovszki *et al.,* 2001). Importantly, these data suggest that methylation of CpG island promoters on the Xi is required to maintain gene repression over time and cell divisions, not to establish the initial heterochromatic state.

Genomic imprinting is a process that results in expression of genes from one chromosome only, i.e. in a parent-of-origin specific manner (reviewed in Ferguson-Smith, 2011). Imprinted genes occur in genomic clusters and can show complex patterns of tissue-specific, ubiquitous and loss of imprinting in different tissues. The mechanisms that lead to monoallelic expression seem to be diverse but DNA methylation has an integral role to play. Regions of the genome can be differentially DNA methylated on one allele relative to the other and this is thought to be required for imprinted expression (Ferguson-Smith, 2011). These differentially methylated regions (DMRs) are established in the germline of each parent and are subsequently maintained in the offspring on their respective alleles. Mouse embryos lacking *Dnmt1* cannot maintain monoallelic expression of imprinted genes and express them from both alleles (Lei *et al.,* 1993). Deletion of *Dnmt3a* in the female germline, the gene encoding the methyltransferase required for establishing imprints, results in loss of imprinting in offspring (Kaneda *et al.,* 2004). Studies of imprinting have provided many insights into epigenetic regulation and have highlighted the diverse ways that DNA methylation can influence gene expression. Multiple genes within an imprinted gene cluster can be regulated simultaneously by the methylation of a single region called an imprinting control region (ICR)(Ferguson-Smith, 2011). These have often been found to be the promoters of non-coding RNAs (ncRNAs) that, when expressed, are thought to establish repressive chromatin domains covering multiple genes by recruiting histone modifying enzymes (Nagano *et al.,* 2008; Ferguson-Smith, 2011). In this way, chromatin domains can

be demarcated by the expression of long, functional ncRNAs that can reportedly act in both *cis* and *trans* (Beard *et al.,* 1995; Rinn *et al.,* 2007). Regulation of ncRNA has been directly linked with DNA methylation at autosomal imprinted genes and *Xist*, the ncRNA required in *cis* for XCI (Beard *et al.,* 1995). It seems likely that this kind of mechanism is more widespread in gene regulation than is currently appreciated and it is interesting to speculate that DNA methylation may be involved in regulating genes in this way outside of imprinted gene clusters. Another role of DNA methylation in gene regulation was established by studying imprinted genes. The DNA binding insulator protein, CTCF, is blocked by CpG methylation that overlaps its binding site and this activity is thought to be essential for the imprinting of *H19*/*Igf2* genes (Bell and Felsenfeld, 2000; Hark *et al.,* 2000). A DMR within the CTCF binding site is methylated on the paternal allele and unmethylated on the maternal allele, resulting in CTCF binding exclusively to the maternal allele. CTCF seems to act by insulating the *Igf2* gene from enhancers present downstream of the *H19* gene, allowing them to act only on *H19*, a situation that is reversed on the paternal allele leading to reciprocal expression of the two genes (Bell and Felsenfeld, 2000; Hark *et al.,* 2000). Despite its potential importance, few further studies have examined the potential role of CTCF exclusion by DNA methylation in regulating non-imprinted genes.

1.2.5.4. Transposable and repetitive element silencing

More than 40% of mammalian genomes are derived from retroelements, many of which are now dormant but a significant proportion are potentially active (Rowe *et al.,* 2010). Retrotransposons can multiply in an individual genome through transcription, reverse transcription to generate double stranded nucleic acid, followed by reinsertion into a new locus (Havecker *et al.,* 2004). It is believed that these elements represent a potential threat to their individual hosts by potentially influencing the expression of nearby genes and inducing mutation by insertion (Karimi *et al.,* 2011; Rebollo *et al.,* 2011). Epigenetic mechanisms are thought to contribute to retrotransposon silencing, including DNA methylation and H3K9me3 mediated heterochromatinisation (Rowe *et al.,* 2010; Karimi *et al.,* 2011; Rebollo *et al.,* 2011). Mammals include DNA methylation of retrotransposons as part of their global methylome but it seems that retrotransposon methylation is evolutionarily ancient and many extant organisms have the bulk of their methylation located at these elements (Lister *et al.,* 2009; Zemach and Zilberman, 2010). It has been speculated that the emergence of a global methylome may have been linked to increases in abundance of retrotransposons in vertebrates and a subsequent need to keep them silent (Zemach and Zilberman, 2010).

Intercisternal A particles (IAPs), a subclass of retrotransposons, are strongly upregulated at the RNA level in *Dnmt1*$^{-/-}$ embryos and various retrotransposons are upregulated in mouse mutants that have defects in DNA methylation although, to my knowledge, an increased retrotransposition has not been reported in these mutants (Walsh *et al.*, 1998; Bourc'his and Bestor, 2004; Huang *et al.*, 2004; De La Fuente *et al.*, 2006; Karimi *et al.*, 2011). The methyl-CpG-binding protein MeCP2 has also been linked to the repression of LINE-1 retrotransposon sequences and their retrotransposition in neural progenitor cells (Muotri *et al.*, 2010). This suggests that the DNA methylation system is required for complete silencing of at least a subset of retrotransposons in mammals and therefore helps to restrict transcription to genic and functional non-coding regions of the genome.

1.2.5.5. DNA methylation of promoter regions of single copy genes

It was suggested over 35 years ago that DNA methylation may act to repress transcription at regulatory regions (Riggs, 1975). Since then, it has become accepted that the activity of gene promoter regions, important regulatory features of genes, can be modulated by DNA methylation. When present, high promoter DNA methylation is inversely correlated with transcription and it is proposed to repress gene expression in many organisms (Weber *et al.*, 2007; Meissner *et al.*, 2008; Lister *et al.*, 2009; Zemach *et al.*, 2010). As discussed in 1.2.4.3, the density of CpG motifs is related to the effect of DNA methylation in a promoter region, an observation that has also been made in artificial systems (Boyes and Bird, 1992). In support of a role for DNA methylation in gene repression, loss of DNA methylation results in activation of gene expression, as demonstrated in many studies. In *Xenopus laevis* embryos, depletion of *xDnmt1* results in premature activation of gene expression (Stancheva and Meehan, 2000). In mouse fibroblasts, mutation of *Dnmt1* reportedly results in activation of ~10% of genes (Jackson-Grusby *et al.*, 2001) and in another study many tissue-specific genes were shown to be activated in *Dnmt1* mutant fibroblasts (Lande-Diner *et al.*, 2007). Mouse ES cells deficient in all three active DNMTs show upregulation of certain developmental genes that would normally be expressed later in development (Fouse *et al.*, 2008; Karimi *et al.*, 2011). Studies that investigate mutants of other *DNMT* genes or DNA methylation co-factors also observe illegitimate gene activation supporting a major role for DNA methylation in gene silencing (Xi *et al.*, 2007a; Ng *et al.*, 2008; Velasco *et al.*, 2010; Borgel *et al.*, 2010).

In mammals, as discussed in 1.2.4, DNA methylation patterns at promoter regions can be variable between tissues. This supports a role for promoter DNA methylation in

repressing lineage-specific genes and hence contributing to an epigenetic memory. One example is at genes involved in maintaining pluripotency in the early embryo and ES cells, that are transcriptionally silenced followed by promoter DNA methylation during differentiation (Feldman *et al.,* 2006; Epsztejn-Litman *et al.,* 2008). Cells lacking *de novo* methyltransferases can silence these genes but are more prone to reactivate them following loss of a differentiation cue (Feldman *et al.,* 2006; Epsztejn-Litman *et al.,* 2008). Another example is a number of genes that are expressed in the germline, where they are unmethylated, and repressed in somatic lineages, where their promoter is methylated (Weber *et al.,* 2007; Borgel *et al.,* 2010; Velasco *et al.,* 2010). In mouse somatic cells and tissues lacking DNA methyltransferases, many of these genes are ectopically expressed (Borgel *et al.,* 2010; Velasco *et al.,* 2010). A link between DNA methylation and germline-expressed genes also comes from studies of cancer, where widespread re-localisation of DNA methylation is observed, including hypomethylation of large intergenic regions and hypermethylation of certain CpG islands (Miranda and Jones, 2007; Hansen *et al.,* 2011). A set of genes that are normally expressed in the germline, sometimes called cancer testes antigens, are upregulated in cancers that are hypomethylated and promoter DNA methylation has been suggested to directly silence these genes in normal tissues (De Smet *et al.,* 1996; De Smet *et al.,* 1999). It is interesting to speculate that many germline genes can be directly silenced by DNA methylation in somatic tissues because they will not lose CpG dinucleotides by deamination over generations. This is because they would be unmethylated and expressed during the germline cycle and methylated and repressed in the somatic tissues, which do not contribute to the next generation. Many CpG islands are aberrantly hypermethylated in a range of cancers (Miranda and Jones, 2007; Hansen *et al.,* 2011). This process is often suggested to contribute to the cancer phenotype as many of these genes can act as tumour suppressors. These observations strengthen the view that DNA methylation acts to repress gene promoters as many of these genes can be activated by demethylation induced chemically or genetically in cancer cells (Miranda and Jones, 2007). However, whether promoter hypermethylation is a cause or consequence of the cancer process is a subject of debate. One study suggests that promoter hypermethylation, in a mouse cancer model of chronic lymphocytic leukaemia, is an early event and occurs before symptoms are observed (Chen *et al.,* 2009). However, DNA methylation may occur at genes that are already transcriptionally inactivated and may not play a role in their initial silencing (Chen *et al.,* 2009; Sproul *et al.,* 2011). In this way, aberrant promoter DNA methylation may act to prevent activation of normally temporarily silent genes and hence could lead to proliferation by preventing activation of genes required for differentiation.

1.2.6. Molecular mechanisms of action

Despite the long standing connection between DNA methylation and transcriptional silencing, the molecular mechanisms by which it functions to repress transcription remain unclear. In this section I will give an overview of the proposed mechanisms and discuss the evidence for each. Current evidence supports a multifactorial effect of DNA methylation, although the best characterised functions can be divided into three broad groups; direct inhibition of sequence-specific DNA-binding proteins, recruitment of proteins that recognise the 5mCpG motif, and blocking the recruitment of proteins that recognise the unmodified CpG motif (Fig. 1.5). I will start by discussing the evidence for these more widely accepted roles of DNA methylation. As many of the proposed mechanisms involve DNA methylation-modulated modification of chromatin, I will discuss this apparently major role in section 1.2.6.4. In addition to the more established roles of DNA methylation, I will discuss evidence for more putative and unexplored roles that have emerged recently.

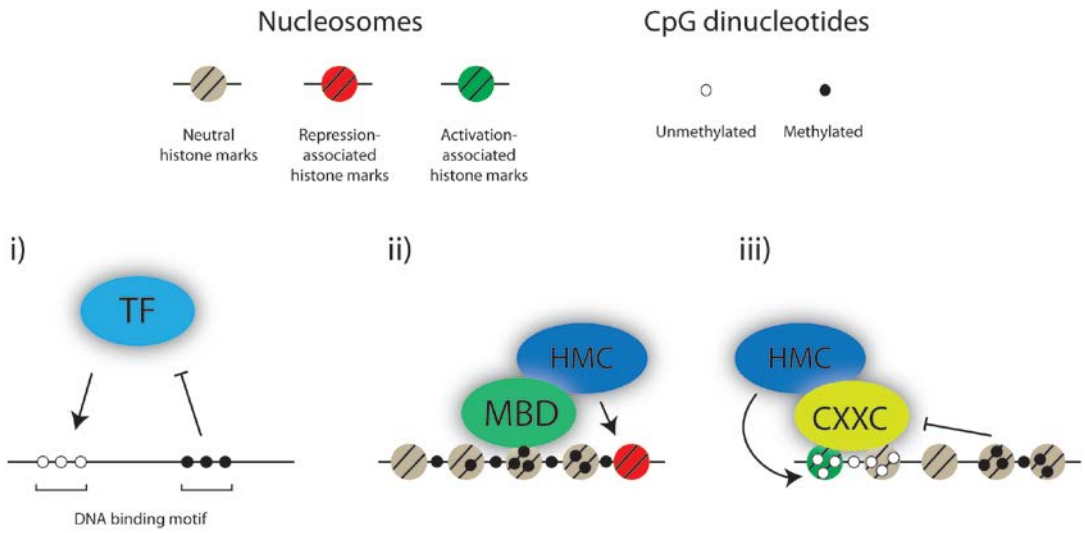


Figure 1.5 - Canonical roles of the 5mCpG dinucleotide. i) 5mCpG can directly inhibit DNA-protein interactions for proteins such as transcription factors (TF) and CTCF. ii) 5mCpG is bound by methyl-CpG-binding proteins, exemplified by a methyl-CpG-binding domain protein (MBD), that can recruit histone modifying complexes (HMC) to induce/maintain a chromatin state. iii) 5mCpG excludes the binding of CXXC-domain containing proteins (CXXC) that recognise unmethylated CpG motifs. Many CXXC-containing proteins have histone modifying activity or are associated with HMCs.

1.2.6.1. Direct inhibition of protein-DNA interactions and effect on DNA structure and mechanics

One of the proposed functions of 5mC in transcriptional regulation is direct inhibition of DNA-protein interactions (Fig. 1.5i). Methylation of cytosine in the CpG context is well known to negatively affect the binding or activity of certain sequence-specific DNA binding proteins, exemplified by restriction endonucleases in bacteria. It is easy to imagine that in this way DNA methylation could have a negative or positive effect on transcription by influencing the binding of transcription factors. A small number of studies have suggested that this type of mechanism is at play in mammalian systems. Watt and Malloy (1988) suggested that site-specific DNA binding of a transcription factor could be inhibited by CpG methylation central to the binding site. Another example is cMyc, which is reportedly inhibited by CpG methylation within its cognate binding site *in vitro* (Prendergast and Ziff, 1991). Methylation of a single CpG has been linked to transcriptional repression at an rDNA gene by preventing the binding of an activating transcription factor (Santoro and Grummt, 2001). In addition, CpG methylation blocks binding of the glucocorticoid receptor to an enhancer region *in vitro* (Wiench *et al.*, 2011). These studies suggest that DNA methylation

may act to suppress transcription by preventing binding of proteins to their cognate DNA sequences, and in this way tissue-specific differences in DNA methylation could contribute to an epigenetic memory. However, the number of factors that have been shown to be influenced by CpG methylation is small, so the extent of the contribution of this mechanism to gene regulation is unclear. This may be in part due to a lack of large scale assays to approach this problem. A recent study developed one such technique and found many new candidates for direct inhibition by CpG methylation, including transcription factors E2F1, HIF1A and components of the TFIIIC complex providing an expanding horizon for this mechanism of transcriptional regulation (Bartke *et al.*, 2010).

Another type of DNA binding protein that can be directly inhibited by CpG methylation is CTCF which binds a core motif that can contain a CpG. As discussed in 1.2.5.3, CTCF exclusion by DNA methylation is thought to control the imprinting of the *H19/Igf2* locus highlighting its capacity for gene regulation (Bell and Felsenfeld, 2000; Hark *et al.*, 2000). CTCF is a protein implicated in chromatin organisation with diverse putative roles in gene expression including insulation and boundary element formation (reviewed in Philips and Corces, 2009). Therefore, the effect of CTCF exclusion by CpG methylation could have complex and widespread effects on gene expression. To date, the more widespread role of DNA methylation in modulating CTCF binding and gene expression at non-imprinted genes is unclear. Some studies have begun to investigate CpG methylation-modulated CTCF binding at non-imprinted genes in normal tissues and cancer and have associated it with gene expression changes (Goto *et al.*, 2009; Lai *et al.*, 2010; Rodriguez *et al.*, 2010; Gao *et al.*, 2011).

The effect of DNA methylation on the mechanical properties of DNA is controversial. A number of studies provide evidence that CpG methylation may decrease the flexibility of DNA in certain contexts (Geahigan *et al.*, 2000; Meints and Drobny, 2001; Nathan and Crothers, 2002). It has been suggested that decreased flexibility could underlie the negative effect of methylation on DNA binding of certain proteins which may need to manipulate flexible DNA. DNA methylation has also been suggested to influence DNA strand separation dynamics, an important step in transcription (Severin *et al.*, 2011). As these studies are performed *in vitro* on naked DNA it is unclear what the effects of DNA methylation are *in vivo* in the presence of enzymes and in the context of nucleosomes. Interestingly, DNA methylation has been linked to increased compaction and rigidity of nucleosome arrays *in vitro* suggesting that DNA methylation *per se* may influence chromatin structure (Choy *et al.*, 2010). In this regard, a study of mouse ES cells lacking DNA methylation showed that some aspects of chromatin organisation were altered, but

compaction of bulk chromatin was unchanged (Gilbert *et al.*, 2007). Further studies in hypomethylated cells and tissues are required to determine the effect of DNA methylation on chromatin structure *in vivo*.

1.2.6.2. Methyl-CpG binding proteins

Another mechanism for transcriptional regulation by DNA methylation is binding by methyl-CpG-binding proteins (MeCPs)(Fig. 1.5ii). Similar to proteins that 'read' histone modifications, it has been suggested that MeCPs play an important role in interpreting 5mCpG. It was first noticed that an unidentified mammalian protein complex bound specifically to methylated DNA, and this complex was termed MeCP1 for 'methyl-CpG-binding protein 1' (Meehan *et al.*, 1989). The first MeCP itself to be identified was subsequently termed MeCP2, and was shown to bind to a single 5mCpG and could repress *in vitro* transcription (Lewis *et al.*, 1992; Meehan *et al.*, 1992). Since then, many MeCPs have been identified and can be split into three groups depending on which methyl-CpG-binding protein domain they possess (Table 1.2). The first methyl-binding protein domain to be identified was termed the MBD (for 'methyl-binding domain')(Nan *et al.*, 1993). Early studies of the structure of the MBD domain bound to 5mCpG suggested that recognition is achieved through a hydrophobic patch on the binding surface of the MBD and that MBD-DNA binding would not be inhibited by core histones (Ohki *et al.*, 2001). Further studies have implicated water molecules in recognition of 5mCpG by the MBD domain of MeCP2, suggesting that MeCP2 recognises hydration of the major groove of methylated DNA rather than the methylated cytosine *per se* (Ho *et al.*, 2008). Humans and mice have four proteins of the MBD family that bind to methylated DNA; MeCP2, Mbd1, Mbd2 and Mbd4 (Table 1.2). Two human proteins with domains similar to the MBD domain were identified, MBD5 and MBD6, but were shown to have no affinity for methylated DNA suggesting that they are not MeCPs (Laget *et al.*, 2010). A further MBD protein, Mbd3, exists in mammals and has important roles in gene silencing but does not bind to methylated DNA (Hendrich *et al.*, 2001). Each of the MBD-containing MeCPs has affinity for symmetrical 5mCpG but some have additional suggested sequence preferences or binding activity. MeCP2 has a binding preference for CpG flanked by runs of AT bases *in vitro* (Klose *et al.*, 2005), Mbd1 can bind to unmethylated DNA through its CXXC domain (Jorgensen *et al.*, 2004), and the structure of the MBD of Mbd2 bound to methylated DNA suggests influence of bases surrounding the 5mCpG (Scarsdale *et al.*, 2011). A second protein domain that can recognise 5mCpG is the modified zinc finger of the Kaiso family of proteins (Table 1.2). Kaiso was the founding

Protein	Associated repressor(s)	Mutant phenotype (mouse)	Reference(s)
MBD domain proteins			
MeCP2	Sin3a complex	Viable and fertile; neurological abnormalities; model of human Rett syndrome (reversible)	Lewis *et al.*, 1992; Guy *et al.*, 2001; Chen *et al.*, 2001; Nan *et al.*, 1998; Jones *et al.*, 1998
Mbd1	Setdb1; Suv39h1; HP1; HDAC3	Viable and fertile; behavioural abnormalities; reduced adult neurogenesis	Cross *et al.*, 1997; Zhao *et al.*, 2003; Sarraf and Stancheva, 2004; Fujita *et al.*, 2003; Villa *et al.*, 2006
Mbd2	NuRD complex; MeCP1 complex	Viable and fertile; defects in maternal behaviour	Hendrich and Bird, 1998; Hendrich *et al.*, 2001; Zhang *et al.*, 1999; Ng *et al.*, 1999
Mbd4	Sin3a complex; HDAC1	Viable and fertile; increased C to T mutation rate; accelerated tumourigenesis on susceptible background	Hendrich and Bird, 1998; Millar *et al.*, 2002; Kondo *et al.*, 2005
MeCP2; Mbd2; Kaiso(Zbtb33)	na	Viable and fertile; neurological abnormalities; similar to *MeCP2* null	Martin Cabellero *et al.*, 2009
Kaiso family proteins (modified zinc finger)			
Kaiso (Zbtb33)	N-Cor	Viable and fertile; decreased tumour formation on susceptible background	Prokhortchouk *et al.*, 2001; Prokhortchouk *et al.*, 2006; Yoon *et al.*, 2003
Zbtb4	Sin3a	Not reported	Filion *et al.*, 2006; Weber *et al.*, 2008
Zbtb38	CtBP	Not reported	Filion *et al.*, 2006; Sasai *et al.*, 2005
SRA domain			
Uhrf1	HDAC1	Early-mid-gestation lethality; similar to *Dnmt1* null	Sharif *et al.*, 2007; Unoki *et al.*, 2004
Uhrf2	Unknown	Not reported	Pichler *et al.*, 2011

Table 1.2 - Associated transcriptional repressors and mouse mutant phenotypes of MeCPs.

member of this family where it was suggested that two 5mCpG motifs in close proximity, preferably immediately adjacent, represent a preferred Kaiso binding site (Prokhortchouk *et al.*, 2001). Subsequently, two proteins with similar zinc fingers have been shown to bind to methylated DNA, Zbtb4 and Zbtb38, and were added to the family (Filion *et al.*, 2006). Binding specificity of Kaiso family proteins has been suggested for both methylated DNA, and non-methylated DNA. Kaiso reportedly binds to CTGCNA motifs and a similar activity

has been attributed to the Kaiso-like proteins (Sasai *et al.,* 2010; Kim *et al.,* 2004). However, the CTGCNA binding activity of Kaiso is not widely conserved and only the methylated DNA binding property of Kaiso is essential for *Xenopus laevis* development suggesting that this is its major function (Ruzov *et al.,* 2004; Ruzov *et al.,* 2009). With the very large number of zinc finger proteins encoded by mammalian genomes it remains likely that more proteins exist that can recognise methylated DNA in this way. A third protein domain that recognises methylated DNA is the SRA domain of Uhrf1 and Uhrf2 (Table 1.2). Uhrf1 reportedly binds hemi-methylated and symmetrically methylated CpG and is important for maintenance methylation involving Dnmt1. The SRA domain of Uhrf1 binds to methylated DNA in a distinct manner from the MBD and zinc finger domains, by flipping out the methylated base (Arita *et al.,* 2008; Avvakumov *et al.,* 2008). Uhrf2 is relatively uncharacterised but contains an SRA domain with 75% similarity to that of Uhrf1 and can bind to hemi-methylated DNA *in vitro* (Pichler *et al.,* 2011).

MeCPs provide a satisfying link between DNA methylation and transcriptional repression as nearly all of them have been suggested to induce transcriptional repression. In concordance with this, all but one of the MeCPs have been found associated with repressor complexes with histone modifying activity (see Table 1.2). In particular, MeCPs are well associated with histone deacetylase (HDAC) activity and the full repressive activity of MeCPs has been suggested to be dependent on this activity in many cases (Table 1.2 refs). An example is MeCP2 which associates with HDACs in the Sin3a complex through its transcription-repression domain (TRD) and inhibition of HDACs by trichostatin A (TSA) relieves MeCP2-mediated silencing (Nan *et al.,* 1998; Jones *et al.,* 1998). MeCP2 is abundant in mouse brain nuclei, at near histone-octomer levels, and is bound broadly to chromatin showing a relatively tight dynamic range that broadly tracks methylated CpG density (Skene *et al.,* 2010). In *MeCP2* null brains increased histone acetylation is observed. This implicates MeCP2 in suppressing transcriptional noise throughout the genome in the brain by reducing histone acetylation. In support of this suggestion, increased Line1 expression and retrotransposition has been reported in *MeCP2*$^{-/-}$ neural progenitor cells but not in fibroblasts (Muotri *et al.,* 2010). MeCP2 may also function in other ways as its interaction with the Sin3a complex is not stable (Klose and Bird, 2004). Also, mapping of HDACs has suggested that they are not stably associated with inactive genes on mammalian chromatin (Wang *et al.,* 2009b). It is of course possible that transient recruitment of HDACs by MeCP2 is sufficient to establish hypoacetylation and that HDACs are not detected due to the transient nature of this interaction with chromatin. MeCP2 has been associated with other chromatin-associated complexes such as SWI/SNF component Brahma and nuclear co-

repressor complex (N-CoR) although some of these observations are controversial and the function of these interactions is currently unclear (Harikrishnan *et al.,* 2005; Kokura *et al.,* 2001; Hu *et al.,* 2006). Another example of a suggested transcriptional repression role of MeCPs is that of Mbd1. Mbd1 is reportedly associated with the histone H3K9 methyltransferase Setdb1 and the H3K9me2/3 binding protein Hp1 (Sarraf and Stancheva, 2004; Fujita *et al.,* 2003). This interaction is proposed to influence the copying of heterochromatin structure during DNA replication. Mbd1 is proposed to form an S-phase specific complex with Setdb1 and Caf-1, a chromatin assembly complex, and facilitate the K9me3 modification of chromatin that is highly DNA methylated (Sarraf and Stancheva, 2004). However, *Mbd1*$^{-/-}$ mice are essentially viable and healthy so it would appear that this mechanism is not essential for heterochromatin assembly on a large scale *in vivo* (Zhao *et al.,* 2003). MeCPs have also been linked to functions outside of transcriptional repression. The major role of Mbd4 is thought to be mismatch repair as *Mbd4* null mice show increased mutation rate at CpG sites (Millar *et al.,* 2002).

Despite the proposed functions of MeCPs in transcriptional silencing, their overall contribution to DNA methylation-mediated transcriptional repression remains unclear. Their role is confounded by studies that show that genetic ablation of individual MeCPs in mice does not lead to overt developmental phenotypes like those of DNA methyltransferase mutant mice (Table 1.2). The exception is *Uhrf1*, the absence of which leads to phenotypes very similar to *Dnmt1* null mice, attributable to its major role in maintenance of DNA methylation (Sharif *et al.,* 2007). Knock-out of *MeCP2* in mice leads to neurological abnormalities that resemble Rett syndrome, a human neurological disorder caused by mutations in *MECP2* (Guy *et al.,* 2001). Intriguingly, re-expression of MeCP2 in null mice in adults rescues the neurological phenotypes indicating that *MeCP2* loss does not lead to irreversible defects in neural development but instead is required for brain function (Guy *et al.,* 2007). The discrepancy between DNA methyltransferase null phenotypes and MeCP null phenotypes suggests that single MeCPs cannot explain transcriptional repression by the DNA methylation system. Instead, data suggests that individual MeCPs contribute, to different extents, to DNA methylation-dependent repression. *Mbd2* deletion results in low level expression of *Xist*, a transcript that is strongly upregulated in a *Dnmt1* mutant, while no mis-expression is observed in cells lacking Mbd1, MeCP2 or Kaiso (Barr *et al.,* 2007). Also, *Mbd2* null cells have a reduced ability to silence methylated reporter plasmids whereas *MeCP2* null cells have only a very small defect in this property (Guy *et al.,* 2001). It is possible that, owing to the number of known MeCPs, functional redundancy between MeCPs masks phenotypes of individual MeCP mutants. This possibility has been partially

investigated by generation of mice lacking three MeCPs; MeCP2, Mbd2 and Kaiso (Martin Cabellero *et al.*, 2009). The phenotype of these mice is very similar to lack of *MeCP2* alone suggesting that these three proteins do not compensate for each other. In order to address the question of redundancy fully, more MeCPs will have to be mutated simultaneously. A further possibility is that a major MeCP is yet to be mutated in mouse. This could include the Zbtb4 or Zbtb38 proteins, which have not yet been knocked out, or as yet unidentified MeCPs. It is of course also possible that MeCPs merely contribute to repression by DNA methylation and that other mechanisms play a role forming a multilayered repression system to ensure that repression is achieved. This idea is supported by the growing number of suggested mechanisms for DNA methylation-mediated repression.

1.2.6.3. Recognition of the unmodified CpG motif

A third mechanism by which DNA methylation is suggested to influence genome regulation is by differentiating CpG islands from bulk chromatin. The CpG dinucleotide alone has been suggested to behave as the shortest known DNA binding motif, as it forms a binding site for proteins with a CXXC domain when unmethylated (Fig. 1.5iii)(Bird, 2011). The CXXC domain is a structure that is capable of binding to DNA with at least one CpG dinucleotide, but importantly binding is inhibited by CpG methylation (Voo *et al.*, 2000; Lee *et al.*, 2001). This type of mechanism could conceivably come under the 'direct inhibition' section but I have decided to discuss it distinctly here and reserve the direct inhibition section for proteins that recognise longer sequence motifs. The structure of the CXXC domain bound to DNA suggested that this binding affinity is achieved through base specific contacts that occur exclusively with the CpG dinucleotide and that methylation of the cytosine likely results in steric hindrance of binding (Cierpicki *et al.*, 2010; Xu *et al.*, 2011a). As most of the genome's CpGs are in the methylated state in vertebrates, a high density of unmodified CpG dinucleotides exists exclusively at CpG islands, differentiating them from bulk chromatin. Recently, the functions of CXXC domain containing proteins in establishing the unique chromatin state at CpG islands have been investigated. Interestingly, many CXXC-domain containing proteins have histone or DNA modifying activity, or are associated with proteins that do, suggesting that targeting chromatin modification may be a major function of these proteins. One such protein, Cfp1 (also known as Cxxc1), is bound to unmethylated CpG islands in mouse cells (Thomson *et al.*, 2010). Cfp1 is associated with the Set1 H3K4me3 methyltransferase complex and is required for normal K4me3 abundance (Thomson *et al.*, 2010). Cfp1 binds and targets K4 methylation to a newly integrated CpG rich region

suggesting that unmethylated CpG dinucleotides are all that is required for its binding. This connection provides a simple explanation for the co-localisation of K4me3 and unmethylated CpG islands. Further study will shed more light on the functions of this protein in epigenetic regulation as it has also been linked to maintaining normal levels of DNA methylation, possibly indirectly through an effect on Dnmt1 protein abundance (Carlone *et al.*, 2005; Butler *et al.*, 2009). Another example of a CXXC protein is Kdm2a, a lysine demethylase that acts on H3K36me2/3 (Blackledge *et al.*, 2010). Kdm2a is recruited to unmethylated CpG islands and is required to keep them depleted of H3K36me2/3. These studies suggest that one of the functions of methylation is to allow CXXC proteins to discriminate unmethylated CpG islands from both methylated CpG islands and bulk chromatin. The effect of this mechanism on gene regulation is currently unknown but one is likely to exist as both the K4me3 and K36me3 have been implicated in the regulation of transcription (reviewed in Kouzarides, 2007). Furthermore, this mechanism is likely to be more widespread as other proteins with CXXC domains that have suggested roles in gene expression exist. Of particular interest, the Mll1 and Mll2/4 proteins contain CXXC domains and are involved in transcriptional activation as part of the trithorax complexes (Terranova *et al.*, 2006). In addition, the Tet1 protein binds to CpG islands in ES cells, presumably through its CXXC domain (Wu *et al.*, 2011). Investigation of other CXXC domain containing proteins will inform us more about how the unique chromatin at CpG islands is established and about this facet of genome regulation by DNA methylation.

1.2.6.4. Maintaining chromatin states

One of the major functions of DNA methylation seems to be in contributing to the establishment and maintenance of chromatin states. This is illustrated by the endpoint of two of the mechanisms described in figure 1.5 being modification of histones. As discussed above, MeCPs and CXXC proteins can link DNA methylation to other modifications of chromatin. Other links have been made between DNA methylation and histone modifications and I will discuss them here.

Eukaryotic cells face a major challenge during DNA replication when they have to dissemble and reassemble chromatin to allow passage of the replication fork, a process that is not well understood (for example Gruss and Sogo, 1992). Correct reassembly of chromatin states is essential for all processes that rely on chromatin structure for a stable memory. Two extremes of chromatin state are often described; euchromatin and heterochromatin, which are broadly characterised by more open, transcriptionally permissive

and more closed, transcriptionally non-permissive chromatin structures respectively, in addition to characteristic histone marks (Beisel and Paro, 2011). How patterns of euchromatin and heterochromatin are maintained over cell division is not entirely clear but DNA methylation, owing to its simple mechanism for propagation during DNA replication, has been implicated in this process. The most convincing model of nucleosome behaviour during replication suggests that nucleosomes from the parental strand are distributed evenly on the two daughter strands following transit of the replication fork (Gruss and Sogo, 1992). This means that with no propagation mechanisms, histone marks would be diluted at each S-phase. DNA methylation may contribute to the propagation of certain histone marks via the Dnmt1 protein itself during its role as maintenance methyltransferase (Fig. 1.6i). Dnmt1 was suggested to interact with HDAC2 and the two proteins co-localise at replication foci during S-phase (Rountree *et al.*, 2000). During S-phase, euchromatic regions tend to replicate early and heterochromatic regions tend to replicate late, a process thought to provide a means for compartmentalising the genome (Rountree *et al.*, 2000). Interestingly, HDAC2 was shown to co-localise with Dnmt1 at replication foci only during late S-phase implicating this mechanism in keeping heterochromatin hypoacetylated. However, this idea is difficult to address experimentally as null mutations in *Dnmt1* become DNA hypomethylated and would therefore be expected to lose MeCP binding, which would be expected to lead to hyperacetylation itself. A subtle mutation in Dnmt1 that would ablate interaction with HDAC2 may shed light on this putative mechanism. In support of a role for Dnmt1 and histone deacetylation in maintaining late replication timing, *Dnmt1* mutant cells reportedly show premature replication of regions of constitutive heterochromatin at pericentric regions (Casas-Delucchi *et al.*, 2011). Dnmt1 also interacts directly with G9a, the histone methyltransferase responsible for the majority of K9me2/3 in euchromatic regions (Esteve *et al.*, 2006). Dnmt1 and G9a co-localise with K9me2 foci during S-phase and Dnmt1 is required for normal G9a localisation to replication foci and K9me2 at target regions (Esteve *et al.*, 2006). This suggests that Dnmt1 recruits G9a to regions of high DNA methylation through its interaction with PCNA, Uhrf1 and hemi-methylated DNA to contribute to chromatin propagation at these regions. These mechanisms implicate Dnmt1 as a 'reader' of DNA methylation patterns by recruiting histone modifying enzymes to DNA methylated regions following DNA replication (Fig. 1.6i). It is possible that propagation of other chromatin states is influenced by Dnmt1 as interactions have been reported with other chromatin-associated proteins such as Suv39h1 and HP1 (Esteve *et al.*, 2006; Smallwood *et al.*, 2007). In this way, the role of the DNA methylation mark can be viewed as a DNA replication-stable template for histone mark propagation by recruiting Dnmt1 and associated

factors following passage of the replication fork (Fig. 1.6i). It should be noted that DNA methylation is not implicated in transmission of all heterochromatin, for example, the H3K27me3 mark is thought to be self-propagated by polycomb proteins (Margueron *et al.*, 2009). Also, organisms that have very low levels of, or completely lack, DNA methylation, such as *Drosophila*, *C. elegans*, budding and fission yeast, can maintain heterochromatin over cell divisions. Fission yeast may have evolved specialised targeting of constitutive heterochromatin involving targeting by small ncRNA, a process that is also linked to DNA replication (Li *et al.*, 2011a). It therefore seems likely that DNA methylation in vertebrates is required to merely contribute to propagation of certain histone states, a role that is perhaps necessary for larger genomes with more complex patterns of heterochromatin.

DNA methylation has also been linked to the deposition of histone variants in chromatin. In *Arabidopsis thaliana*, the locations of DNA methylation and the histone variant H2A.Z are negatively-correlated (Zilberman *et al.*, 2008). In addition, these epigenetic components appear to be mutually exclusive as DNA methylation changes induced by mutation of the plant maintenance methyltransferase, *MET1*, lead to reciprocal changes in H2A.Z location, and loss of H2A.Z leads to hypermethylation (Zilberman *et al.*, 2008). The anti-correlation between DNA methylation and H2A.Z is thought to be evolutionarily ancient as, in addition to plants, it has been observed in animals. Mapping of pufferfish DNA methylation and H2A.Z showed a very striking negative-correlation (Zemach *et al.*, 2010). In mammals, changes in H2A.Z and DNA methylation during oncogenesis are inversely correlated implying a similar relationship, and acetylation of H2A.Z is also negatively-correlated with DNA methylation (Valdes-Mora *et al.*, 2011; Conerly *et al.*, 2010). The function of H2A.Z in gene regulation is controversial but among other things it has been linked to maintaining activity of expressed genes, perhaps through its propensity to dissociate from chromatin leaving a nucleosome depleted region (Zhang *et al.*, 2005). As unmethylated CpG islands have been linked to nucleosome destabilisation, it is tempting to speculate that a function of DNA methylation is to prevent the deposition of H2A.Z, thereby promoting a higher density of nucleosomes (Tazi and Bird, 1990; Ramirez-Carrozzi *et al.*, 2009). It should be noted, however, that it is currently unclear if CpG islands are intrinsically nucleosome destabilising or if this association arises because vertebrates often have a nucleosome depleted region at gene promoter regions (Deaton and Bird, 2011). Further study of the functions of H2A.Z and the molecular mechanisms behind its mutual antagonism with DNA methylation are needed to shed light on this interesting putative function of DNA methylation.

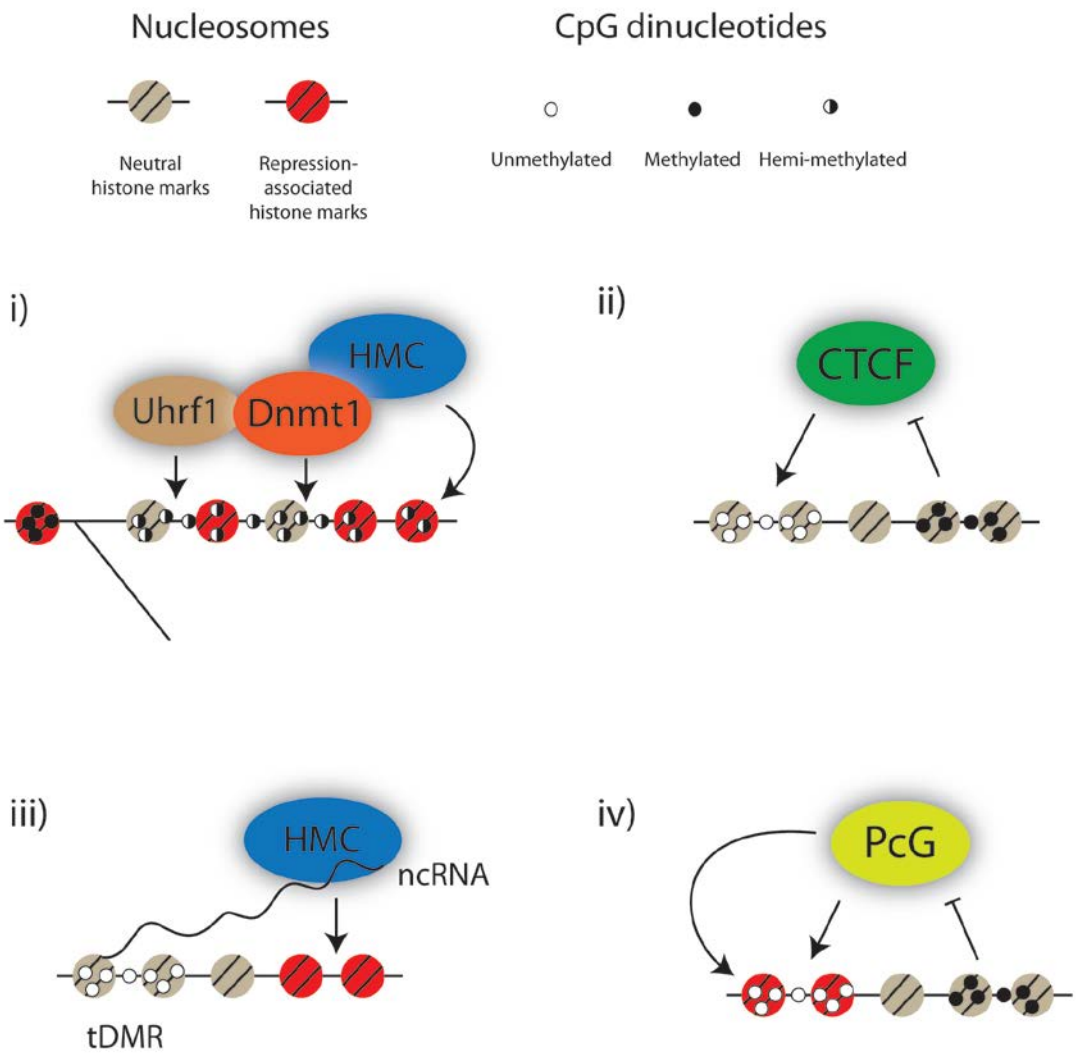


Figure 1.6 - Putative roles of DNA methylation in chromatin structure and gene regulation. i) Dnmt1 may act a 'methylation reader' by recruiting histone modifying complexes (HMC) during maintenance methylation. Nucleosomes are thought to be shared between the two daughter strands after DNA replication leading to a dilution of histone marks. DNA methylation may provide a template for re-establishing a certain chromatin state after DNA replication by recruiting Dnmt1/Uhrf1 and associated proteins to hemi-methylated DNA. ii) Tissue-specific DNA methylation at CTCF binding sites may lead to differential binding and complex effects on the expression of genes in the region. Similar to the imprinted *H19/Igf2* locus. iii) Methylation at a tDMR (tissue-specific differentially methylated region) could directly control the expression of ncRNAs with roles in regulating gene expression in *cis* or *trans*. ncRNAs may contribute to initiation/maintenance of a chromatin state by recruiting histone modifying complexes (HMCs) leading to complex effects on local gene expression. Similar to the *Xist* ncRNA in XCI. iv) DNA methylation may inhibit the binding of polycomb group proteins (PcG) to chromatin, directly or indirectly, contributing to their distribution. As PcG proteins are involved in gene repression, in this way DNA methylation may lead to gene activation.

1.2.6.5. Other roles for DNA methylation?

Recent studies are providing intriguing evidence that the effects of DNA methylation are likely more numerous and complex than currently appreciated (examples of putative mechanisms are shown in fig. 1.6ii-iv). A recent study used *in vitro* assembled nucleosomes with either methylated or unmethylated biotinylated DNA coupled with SILAC (Stable Isotope Labelling with Amino acids in Cell culture) to 'fish' for proteins that are differentially bound by methylated/unmethylated DNA on nucleosomes (Bartke *et al.*, 2010). Using only two types of DNA, excluding the possibility of finding many sequence-specific DNA binding factors, they were able to identify many new candidates for methylation-modulated binding. Proteins that preferentially bound methylated DNA included zinc finger protein ZHX1, and transcription factors FOXA1, PAX6 and HOXC10. Negatively affected proteins included transcription factors in addition to chromatin associated proteins such as the PRC2 complex, BCOR complex and CXXC-containing CXXC5 providing new candidates for direct inhibition. These proteins must be viewed as candidates for methylation-regulated binding to chromatin as these experiments do not infer whether the effect is direct or indirect. Further studies using this technique may use a wider variety of DNA sequences and combinations of histone marks and DNA methylation to provide important information about cross-talk between epigenetic modifications of chromatin.

It will be interesting to learn if mechanisms of DNA methylation-mediated regulation at imprinted genes are also applicable to the rest of the genome. For example, blocking of CTCF binding by DNA methylation (as discussed in 1.2.6.1). Mapping of CTCF in mammalian cell lines has shown that although many CTCF binding sites are conserved between cell types, a proportion appear to cell-type specific (Kim *et al.*, 2007). In addition, a large proportion of CTCF binding sites contain a CpG dinucleotide (Kim *et al.*, 2007). This suggests the possibility that differential DNA methylation at these sites could control CTCF binding in a tissue-specific manner leading to potentially complex effects on regulation of the genome (Fig. 1.6ii). Some studies have begun to test this idea, including a study of the *Pax6* locus where CTCF binding was shown to be anti-correlated with DNA methylation throughout ES cell differentiation (Gao *et al.*, 2011). With the advent of base-resolution methylation maps and detailed CTCF binding profiles it will be interesting to test this hypothesis on a larger scale. Another concept that has emerged from imprinted genes is DNA methylation controlling a locus indirectly by regulating the expression of a functional ncRNA. *Xist* is an example of a lincRNA (long intergenic non-coding RNA) that has long been known to be regulated by DNA methylation and is responsible for *in cis* inactivation of

large regions of one X-chromosome, a process that may involve direct targeting of histone modifying complexes (Beard *et al.,* 2005; Augui *et al.,* 2011). LincRNAs are numerous in mammals, have been implicated in both activation and repression of gene expression by association with HMCs and can reportedly act in both *cis* and *trans* (Rinn *et al.,* 2007; Wang *et al.,* 2011; Cabili *et al.,* 2011). It remains untested if DNA methylation regulates ncRNAs in a tissue-specific manner, out with of imprinted genes, that can influence chromatin structure either in cis or trans. One possibility is that 'orphan' CpG islands (CGIs) represent promoters for some of these functional ncRNAs. Orphan CGIs are methylated more frequently in a tissue-specific manner than gene-associated CGIs and many show hints of promoter function (Illingworth *et al.,* 2008; Illingworth *et al.,* 2010). A recent study showed lincRNA expression to be highly tissue-specific in keeping with this speculation (Cabili *et al.,* 2011).

A handful of studies have suggested a putative role for DNA methylation in the regulation of the H3K27me3 mark mediated by the PRC2 (polycomb repressor complex 2) complex (Fig. 1.6iv). This will be discussed in the next section, following a brief introduction to the polycomb group proteins.

1.3. Polycomb group proteins

Another epigenetic system that has a major function in gene regulation in multicellular organisms is the polycomb repression system. Polycomb group (PcG) and trithorax group (TrxG) proteins form multimeric complexes with chromatin modifying activity and are thought to act antagonistically with each other to coordinate transcriptional memory (Cavalli and Paro, 1999; Czermin *et al.,* 2002; Boyer *et al.,* 2006; Bracken *et al.,* 2006; reviewed in Schuettengruber *et al.,* 2007). PcG repression fulfils the strictest definition of an 'epigenetic mechanism' as it is transmitted through cell divisions (Hansen *et al.,* 2008). These protein complexes have been implicated in many interesting biological processes including gene regulation, lineage specification, genomic imprinting, stem cell identity and cancer, provoking intense recent study of their functions. In this section I describe PcG proteins and their role in epigenetic regulation of gene expression. This section is less detailed than section 1.2 as DNA methylation was the initial focus of this thesis project.

PcG and TrxG proteins are thought to have opposing roles in gene regulation; PcG proteins are involved in maintaining lineage-specific gene repression while TrxG proteins are involved in maintaining lineage-specific gene expression (Cavalli and Paro, 1999; Boyer *et al.,* 2006; Bracken *et al.,* 2006; reviewed in Schuettengruber *et al.,* 2007). A key suggestion is that these proteins are required specifically to maintain the expression status of their target genes, not to initiate patterns of expression (Cavalli and Paro, 1999). This strongly implicates these complexes in forming part of the epigenetic memory of gene expression states that is thought to be essential for development of complex multi-cellular organisms. Most PcG and TrxG genes were originally identified in *Drosophila* as mutations that cause or modify homeotic phenotypes; body axis translocations caused by mis-expression of homeotic (*Hox*) genes (reviewed in Schuettengruber *et al.,* 2007). The next section will describe the known PcG and TrxG proteins and protein complexes in fly and their orthologs in mammals.

1.3.1. PcG and TrxG proteins in fly and man

PcG and TrxG proteins are well conserved throughout eukaryotes (reviewed in Schuettengruber *et al.,* 2007). PcG genes are found in a wide variety of multi-cellular eukaryotes, and even some single-celled eukaryotes, but are noticeably absent from the

single celled organisms *S. pombe* and *S. cerevisiae* (Schuettengruber *et al.*, 2007). The protein products of PcG genes form complexes. In *Drosophila* at least four major PcG protein complexes exist; PRC2 (polycomb repressor complex 2), PRC1, PhoRC (Pho-repressive complex) and PR-DUB (Polycomb repressive de-ubiquitinase) complexes, of which PRC2 and PRC1 have been most studied (Schuettengruber *et al.*, 2007; Scheuermann *et al.*, 2010). The components and formation of the PRC2 and PRC1 complex are well conserved in humans. A PR-DUB complex has been reported in humans with similar activity suggesting it too is conserved (Scheuermann *et al.*, 2010; Bott *et al.*, 2011). Human homologs of PhoRC components have been suggested but their contribution to PcG protein-complexes and functions are not clear (Schuettengruber *et al.*, 2007). Fly PRC2 and PRC1 core complex components are shown in table 1.3 together with their human orthologs and suggested functions. The *Drosophila* PRC2 complex consists of four core components; E(z) (enhancer of zeste), Esc (extra sex combs), Su(z)12 (suppressor of zeste 12) and N55 (table 1.3). In mammals, the PRC2 and PRC1 core components are conserved but there are multiple possible members of certain PRC1 components that are thought to have arisen through gene duplication. For example, fly Pc has at least five homologs in human, giving multiple possible configurations for the human PRC1 complex and adding to the complexity. Other proteins have been associated with the PRC2 and PRC1 complexes but will not be discussed here (Schwarz and Pirrotta, 2007; Margueron and Reinberg, 2011). Fly TrxG complexes include the nucleosome remodelling SWI/SNF complex and proteins with histone modifying activity; Trx has methyltransferase activity for H3K4 and Ash1 has reported methyltransferase activity for multiple histone sites (Schwarz and Pirrotta, 2007). The human homologs of *Drosophila* Trx are the MLL proteins and although they seem to be functionally homologous to the fly protein, they do not form the same complexes (Schwarz and Pirrotta, 2007). Herein, for simplicity I will refer to polycomb and trithorax proteins by their names in mice, which are the same as human.

Both PRC2 and PRC1 harbour histone modifying activity that is bestowed by their Ezh2/Ezh1 and Ring1b/Ring1a subunits respectively (Cao *et al.*, 2002; Czermin *et al.*, 2002; de Napoles *et al.*, 2004; Wang *et al.*, 2004). A general model for PcG function is the sequential recruitment model, whereby PRC2 is targeted to a chromosomal region, catalyses H3K27me3, which forms a binding site for the PRC1 complex through its CBX component that possesses a chromodomain (Cao *et al.*, 2002; Wang *et al.*, 2004b). This model is supported by the enrichment of PRC1 binding to genomic regions where H3K27me3 is abundant (Cao *et al.*, 2002). However, PRC2 and PRC1 are at least partially independent as PRC1 is found at many locations without H3K27me3 and the two complexes have many

non-overlapping gene targets for repression (Leeb *et al.*, 2010; Ku *et al.*, 2008). Both PRC2 and PRC1 complexes play important roles during embryogenesis as mutation of polycomb genes in mouse leads to a range of phenotypes, from early embryonic lethality in PRC2 components, to developmental arrest during gastrulation or skeletal abnormalities for PRC1 components (Voncken *et al.*, 2003; O'Carrol *et al.*, 2001; van der Lugt *et al.*, 1994). The molecular basis for these phenotypes is largely attributed to the fact that many important genes are targeted for repression by PcG proteins during development. The major genomic targets of PcG proteins will be discussed in the next section.

Complex	Fly	Human	Function	Reference(s)
PcG complexes				
PRC2	E(z)	EZH2; EZH1	H3K27me3 deposition	Cao *et al.*, 2002; Czermin *et al.*, 2002
	Esc	EED	Binding/propagation of H3K27me3	Margueron *et al.*, 2009
	Su(z)12	SUZ12	PRC2 activity	Cao and Zhang, 2004
	N55	RpAp48; RpAp46	Histone binding	Song *et al.*, 2008
PRC1	dRing	RING1A; RING1B	H2A K119ub deposition	de Napoles *et al.*, 2004; Wang *et al.*, 2004a
	Pc	CBX2; CBX4; CBX6; CBX7; CBX8	Chromatin binding; K27me3 binding	Cao *et al.*, 2002
	Ph	PHC1; PHC2; PHC3	Unknown but required for PRC1 function	Isono *et al.*, 2005
	Psc	BMI1; MEL18; NSPC1	Co-factor for dRing	Cao *et al.*, 2005; Elderkin *et al.*, 2007; Wu *et al.*, 2008

Table 1.3 - PRC2 and PRC1 core complex components in fly and man. Only the core components are shown. The fly protein is shown with its homologous human proteins and suggested function. Adapted from Schwarz and Pirrotta (2007) and Schuettengruber *et al.,* (2007).

1.3.2. PRC2 targets and the H3K27me3 mark

PRC2 component mapping in mammalian cells has shown that this complex has a strong enrichment for binding at genes with roles in embryonic development (Bracken *et al.*, 2006; Lee *et al.*, 2006; Boyer *et al.*, 2006; Schuettengruber *et al.*, 2009). PRC2 marks many tissue-specific genes that encode transcription factors involved in lineage commitment, such as the *Hox* (homeobox) genes. The H3K27me3 (K27me3) histone mark is a signature of the PRC2

complex as Ezh2, and to a lesser extent Ezh1, are the only endogenous proteins known to deposit this mark *in vivo* (Czermin *et al.*, 2002; Margueron and Reinberg, 2011). Therefore, K27me3 is often used as a surrogate for mapping PRC2-repressed chromatin. Globally, K27me3 is enriched at gene promoters and those marked by K27me3 tend to be low or silent in expression implicating it in gene repression (Mikkelsen *et al.*, 2007; Barski *et al.*, 2007). In keeping with this idea, mutation of PRC2 components results in loss of K27me3 and upregulation of lineage-specific genes in ES cells (Cao and Zhang, 2004; Boyer *et al.*, 2006). In addition to marking individual gene promoters, K27me3 can form large blocks of modified chromatin (Lee *et al.*, 2006; Barski *et al.*, 2007; Pauler *et al.*, 2009). Blocks of K27me3 are sometimes referred to as facultative heterochromatin, in contrast to constitutive heterochromatin that is characterised by H3K9me3 and H4K20me3 (Margueron and Reinberg, 2011; Beisel and Paro, 2011). K27me3 blocks are found in gene-rich regions of chromatin, interspersed with gene-poor blocks of H3K9me3 and H4K20me3 (Pauler *et al.*, 2009). This difference is illustrated by the α- and β-globin genes. In non-expressing cell types the α-globin genes, which are located in a gene-dense region, are marked and repressed by K27me3, while the β-globin genes, which are located in a gene-poor region, are not and are silenced by other mechanisms (Garrick *et al.*, 2008). Recently, K27me3 was also found enriched at enhancers (Rada-Iglesias *et al.*, 2011). Specifically, K27me3 marks 'poised' enhancers in ES cells that would become active later in development implicating it in the regulation of distal regulatory elements. A key aspect of repression by K27me3 is that it is flexible and so is suited to regulating genes involved in lineage commitment. Upon lineage commitment, many genes show dynamic patterns of K27me3, including gains and losses, which are reflected by changes in their expression status (Mikkelsen *et al.*, 2007; Mohn *et al.*, 2008). In ES cells, before lineage commitment is initiated, a large number of developmental gene promoters are marked by both the transcription initiation-associated H3K4me3 (K4me3) and the polycomb-associated K27me3 histone marks, together termed the 'bivalent mark' (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007; Mohn *et al.*, 2008). This chromatin configuration has been associated with a 'poised' RNA polymerase II that carries phosphorylation at serine 5 of its C-terminal domain and may be restrained from elongation by polycomb proteins through their modification of chromatin (Stock *et al.*, 2007). Upon differentiation, many of these promoters resolve to either K4me3 or K27me3 marks alone in a lineage-specific manner, although some remain bivalently marked (Mikkelsen *et al.*, 2007; Mohn *et al.*, 2008). Although the bivalent mark is also present in zebrafish, it is very rare in the *Drosophila* embryo suggesting that this feature of PcG/TrxG regulation is not always conserved (Schuettengruber *et al.*, 2009; Vastenhouw *et al.*, 2010).

1.3.2.1. Hox *genes*

Hox genes, encoding homeobox transcription factors, are major targets of regulation by PcG and TrxG proteins and are often used as a model to study repression by polycomb (for example Cao *et al.*, 2002; Cao *et al.*, 2005). As mentioned previously, study of phenotypes in fly that result in mis-regulation of *Hox* genes resulted in the discovery of PcG and TrxG proteins. *Hox* genes encode sequence-specific transcription factors that regulate the formation of a body plan during development (reviewed in Soshnikova and Duboule, 2009a). In particular, *Hox* genes are required to establish and maintain body axes like the anterior-posterior axis of the embryo and during digit development. *Hox* genes have proved interesting from an evolution of body plan perspective and also in regard to gene regulation (Soshnikova and Duboule, 2009a). *Hox* genes are arranged into genomic clusters in many organisms, and in mammals they exist in four clusters on different chromosomes, termed *HoxA-D* clusters. Two fascinating aspects of *Hox* regulation are their temporal and spatial expression patterns. Firstly, *Hox* genes are expressed in the early embryo in the order that they are positioned within their cluster, a behaviour termed temporal co-linearity (Soshnikova and Duboule, 2009b). Temporal co-linearity is thought to involve the clustered arrangement of *Hox* genes as histone marks have been observed to spread across the cluster corresponding with expression changes, and a mouse with a split *HoxD* cluster shows defects in this process (Soshnikova and Duboule, 2009b). Spatially, *Hox* genes are also expressed in a pattern according to their position within their cluster during development, termed spatial co-linearity (Soshnikova and Duboule, 2009a). The clustered arrangement of *Hox* genes does not seem to be essential for this process as *Hox* transgenes are expressed in the correct tissues, but PcG proteins are thought to be important (Simon *et al.*, 1992; Schwarz and Pirrotta, 2007). Some elements of this spatial co-linearity are preserved in adult cells where it is suggested that a pattern of *Hox* expression maintains tissue identity (Rinn *et al.*, 2009; Wang *et al.*, 2011). *Hox* genes are also tightly regulated and required for lineage commitment of adult multipotent cells and are strongly linked to differentiation decisions, for example in haematopoiesis (He *et al.*, 2011b). Together, these features suggest that a '*Hox* code' is important to maintain a sense of positional or lineage identity and that regulation of this code is crucial in cell differentiation.

PcG and TrxG proteins are strongly linked to *Hox* regulation, and maintain repression and activation of *Hox* genes respectively (Simon *et al.*, 1992; Boyer *et al.*, 2006; Wang *et al.*, 2009c). The K27me3 modification marks a large domain encompassing the

individual *Hox* clusters in undifferentiated mouse ES cells whilst their promoter regions, among other punctate regions, are marked by H3K4me3 (Boyer *et al.*, 2006; Lee *et al.*, 2006; Soshnikova and Duboule, 2009b). In differentiated tissue, domains of K27me3 and H3K4me3 within a cluster often demarcate the expression of *Hox* genes, with low K27me3 and high H3K4me3 marking regions of higher active expression (Soshnikova and Duboule, 2009b; Rinn *et al.*, 2007; Wang *et al.*, 2011). Upregulation of *Hox* genes is observed in a range of tissues and cell types when polycomb proteins are mutated, suggesting that the PcG system is critical for *Hox* gene regulation (Simon *et al.*, 1992; Cao and Zhang, 2004; Cao *et al.*, 2005; Boyer *et al.*, 2006; Fujimura *et al.*, 2006; Stock *et al.*, 2007; Chamberlain *et al.*, 2008; Eskeland *et al.*, 2010). Defects in expression of *Hox* genes are also observed in mutants of MLL proteins (Terranova *et al.*, 2006; Wang *et al.*, 2009c). In particular, Mll1 is required for normal levels of H3K4me3, RNA Pol II (Pol II) targeting and expression of *Hox* genes in mouse fibroblasts, supporting the opposing roles of PcG and TrxG proteins in gene regulation (Wang *et al.*, 2009c).

1.3.3. Mechanisms of repression by PcG complexes

PcG proteins are well associated with maintaining gene repression but the mechanisms by which they achieve this are unclear. Here I will briefly discuss the multiple mechanisms that have been suggested recently. The sequential recruitment model predicts that PRC2 inhibits transcription by recruiting PRC1 (Wang *et al.*, 2004). However, as discussed above this may not always be the case as PRC1 and PRC2 binding are often non-overlapping (Ku *et al.*, 2008). In addition, many organisms, such as *A. thaliana* and *C. elegans*, have PRC2 components and K27me3 but no K27me3 binding Cbx ortholog to interpret it, suggesting that PRC2 may contribute to gene regulation through PRC1-indepenedent mechanisms (Schuettengruber *et al.*, 2007). PcG proteins have been suggested to inhibit transcript elongation by Pol II. Many PcG target genes are bound by PolII when repressed suggesting that polycomb does not act to inhibit Pol II initiation (Stock *et al.*, 2007). Instead, it has been suggested that PRC1, potentially through its H2AK119 ubiquitin-transferase activity, acts to inhibit Pol II elongation (Stock *et al.*, 2007). This model is yet to be tested biochemically. Secondly, PcG systems may act to inhibit the formation of a more active chromatin state. A long-standing idea is that PcG and TrxG binding to chromatin are mutually exclusive, thus one action of PcG silencing may be to prevent TrxG-mediated transcriptional activation (reviewed in Schuettengruber *et al.*, 2007). However, the relationship between PcG and TrxG proteins is likely to be more complex, as bivalently

marked chromatin is present in mammals and PcG and TrxG proteins physically interact and are co-bound at many sites in *Drosophila* (Schuettengruber *et al.,* 2009; Strubbe *et al.,* 2011). PRC2 is also implicated in histone deacetylation. The PRC2 component Eed interacts with HDACs, and this is thought to contribute to its repressive activity (van der Vlag and Otte, 1999). The histone H3 tail can be acetylated at lysine 27 (K27ac), the position of tri-methylation mediated by PRC2. K27ac and K27me3 are mutually exclusive modifications and K27ac has been linked to transcriptional activity (Tie *et al.,* 2009; Pasini *et al.,* 2010b). Thus, one of the functions of PRC2 could be to induce and maintain hypoacetylation at its target chromatin. Thirdly, PcG proteins have been linked to the organisation of chromatin structure, such as the extent of chromatin compaction. Loss of Ring1b in mouse ES cells results in decompaction of chromatin and transcriptional upregulation of *Hox* genes (Eskeland *et al.,* 2010). These effects are rescued by a catalytically inactive Ring1b, suggesting that the H2AK119Ub mark is not required for chromatin compaction at this locus (Eskeland *et al.,* 2010). Furthermore, it has been suggested that PcG proteins form foci in the nucleus, termed PcG bodies, suggesting a higher level of nuclear organisation of their target chromatin (Hodgson and Brock, 2011). This idea has been recently expanded by the observation that PcG-bound genes form interactions with other PcG-bound genes both in *cis* and in *trans* (Bantignies *et al.,* 2011; Tolhuis *et al.,* 2011). Mutation of a PcG binding site has a subtle negative effect on these interactions suggesting that polycomb proteins may play a role in this level of nuclear organisation (Bantignies *et al.,* 2011). The formation of higher order chromatin structures at PcG targets may be linked to insulator elements which are suggested to form chromosomal contacts as part of their function (Comet *et al.,* 2011; Li *et al.,* 2011b). Intriguingly in this regard, the PRC2 component Suz12 has been shown to interact with CTCF, a protein implicated in chromatin architecture and chromatin interactions (Li *et al.,* 2008; Philips and Corces, 2009). Also, cohesin components have been shown to associate with PRC2 and are required for polycomb-mediated repression of reporters (Strubbe *et al.,* 2011). However, currently the cause and consequence relationships between PcG repression and nuclear architecture are unclear but it is interesting to speculate that PcG bodies act as nuclear repression domains, in the same way that transcription factories act as gene expression domains (Hodgson and Brock, 2011).

1.3.4. PRC2 targeting and maintenance of K27me3

How are PcG complexes targeted? This aspect of the polycomb system is not well understood and represents a major challenge in the field. None of the PRC2 or PRC1 core complex components in table 1.3 have sequence-specific DNA binding activity that is reported, although some interacting proteins may have such a property (Schuettengruber and Cavalli, 2009). In *Drosophila*, DNA sequence elements that are thought to contribute to polycomb recruitment have been identified (Figure 1.7i)(Cavalli and Paro, 1998; reviewed in Schuettengruber and Cavalli, 2009). These elements, termed polycomb response elements (PREs), are thought to be bound by sequence-specific DNA-binding proteins, such as Pho and Phol (Pho-like), that recruit PcG proteins. However, the extent of PcG targeting in this way is unclear as mutation of these DNA-binding proteins in fly does not lead to obvious PcG phenotypes, and no single factor binding site is capable of recruiting PcG complexes (Schuettengruber and Cavalli, 2009). In mammals, the search for PREs has produced few results so far, although a couple of PRE-like elements have been described (Sing *et al.,* 2009; Woo *et al.,* 2010). These elements are reported to recruit PcG-repression in the tissues tested and both implicated the mammalian homolog of Pho, YY1, in this process (Sing *et al.,* 2009; Woo *et al.,* 2010). In addition, mammalian YY1 can act to silence PcG targets in a *Drosophila* reporter system making it an attractive candidate for vertebrate polycomb targeting factor (Atchison *et al.,* 2003). However, studies of YY1 binding in mammalian cells show that it binds to many non-PcG target genes (Xi *et al.,* 2007b). In fact, paradoxically, YY1 binding motifs are enriched in CpG islands that are not bound by PRC2 complex components in mouse ES cells (Ku *et al.,* 2008). Other accessory proteins have been implicated in targeting of the PRC2 complex through recruitment, such as Jarid2 and fly Pcl2 (polycomb-like 2)(Pasini *et al.,* 2010a; Casanova *et al.,* 2011). In the search for DNA sequence motifs at mammalian PcG binding sites, no clear enrichments have been found, but a link has been made with CpG islands (Figure 1.7i). It was noticed that the PcG-bound α-globin locus contains CpG islands while the non-PcG-bound β-globin locus does not (Garrick *et al.,* 2008). This correlation is valid on a larger scale as PcG binding sites are very often found in CpG islands with high levels of sequence conservation (Tanay *et al.,* 2007; Ku *et al.,* 2008; Mendenhall *et al.,* 2010). This idea was functionally tested by inserting CpG-rich DNA from a variety of sources into the mammalian genome. CpG-rich regions that were depleted of activating transcription factor binding sites accrued PcG, suggesting that the default state of CpG-rich DNA may be PcG marking in ES cells (Mendenhall *et al.,* 2010; Lynch *et al.,* 2011). Although Ezh2 binding maps to CG-rich

regions, the K27me3 modification reportedly shows a broader distribution and is in fact depleted from CGI island chromatin (Barski *et al.,* 2007; Mendenhall *et al.,* 2010; Thomson *et al.*, 2010). Instead, K27me3 appears to be enriched in regions flanking CGIs suggesting a 'spreading' mechanism from PRC2 binding sites (Mikkelsen *et al.,* 2007; Schuettengruber *et al.,* 2009). A major role for ncRNA in PcG targeting is emerging, providing another potential source of sequence-specificity. *Xist*, the ncRNA required for XCI, contains a PRC2 binding secondary structure and is thought to recruit PRC2 to the X chromosome that is to be inactivated (Zhao *et al.,* 2008). An RNA-based targeting mechanism looks to be more widespread as PRC2 associates with a large number of ncRNAs (Figure 1.7i)(Zhao *et al.,* 2010; Kanhere *et al.,* 2010). Many polycomb-bound genes produce small RNAs from their 5' end that have been implicated in PRC2 targeting (Kanhere *et al.,* 2010). An RNA-based targeting system has been proposed to function in both *cis*, for example at the 5' end of genes, and *trans*, such as the *HOTAIR* ncRNA that reportedly recruits PcG to the *HoxD* locus despite being expressed from the *HoxC* cluster (Rinn *et al.,* 2007). However, the importance of the homologous *Hotair* transcript in mouse is in question as its absence does not lead to obvious phenotypes usually associated with *Hox* gene mis-expression (Schorderet and Duboule, 2011). Further study of these mechanisms is required to firmly establish a role for RNA-based targeting of PcG proteins. Non-coding RNA expression has also been associated with negative regulation of PcG binding and this will be discussed in section 1.3.5.

In order to provide a component of transcriptional memory, repression by PcG proteins must be transmitted through cell divisions, particularly S-phase when chromatin is disassembled. Recently, it has been suggested that this property is fulfilled by the ability of the PRC2 complex to bind its own mark, K27me3 (Hansen *et al.,* 2008; Margueron *et al.,* 2009). This model suggests that following DNA-replication, PRC2 is recruited to chromatin by existing K27me3, which is directly bound by the C-terminus of Eed (Margueron *et al.,* 2009). This binding allosterically activates the catalytic function of PRC2 leading to deposition of K27me3 on neighbouring newly synthesised histones that are incorporated into chromatin. This form of propagation may also explain the spreading of the K27me3 mark observed during ES cell differentiation (Hawkins *et al.,* 2010). PRC1 and MLL components have been observed to remain chromatin bound during *in vitro* DNA replication and M-phase respectively, suggesting that they are not necessarily displaced from chromatin during these periods (Francis *et al.,* 2009; Blobel *et al.,* 2009).

1.3.5. Relief from polycomb-mediated silencing and polycomb inhibition

One aspect of PcG biology that is important for the work in this thesis is how the flexibility of repression by PcG proteins is achieved. How are PcG-repressed genes activated, and, what is inhibitory to PcG repression? These questions are also relevant to the above section on how PcG proteins are targeted as they involve ways that the cell protects a region from PcG repression. Recently, multiple mechanisms have been suggested in answer to these questions. Firstly, the histone marks associated with the PRC2 and PRC1 complexes are reversible (Figure 1.7ii). UTX and JMJD3 are K27me3 demethylases implicated in reversing PcG-mediated repression and activation of *Hox* genes (Agger *et al.*, 2007; Lan *et al.*, 2007; Seenundun *et al.*, 2010). UTX interacts with MLL2/4 providing a potential explanation for PcG and TrxG antagonism (Issaeva *et al.*, 2007). It is not yet clear which mechanisms govern removal of K27me3 from regulatory regions. Interestingly, a distant enhancer at the α-globin locus has been implicated in recruiting the K27me3 demethylase JMJD3 to a gene promoter, as deletion of the enhancer results in persistence of promoter K27me3 (Vernimmen *et al.*, 2011). The H2AK119 ubiquitination mark catalysed by PRC1 is also thought to be reversible, via the de-ubiquitinase PR-DUB complex (Scheuermann *et al.*, 2010). Paradoxically, mutation of PR-DUB results in the expected increase in H2AK119ub but concomitant derepression of PcG target genes, suggesting a more complex role for this modification in gene regulation (Scheuermann *et al.*, 2010). Also in support of a role for H2AK119ub in gene activation, ZRF1 was identified as a protein that binds to this mark and is involved in transcriptional activation of PcG targets (Richly *et al.*, 2010).

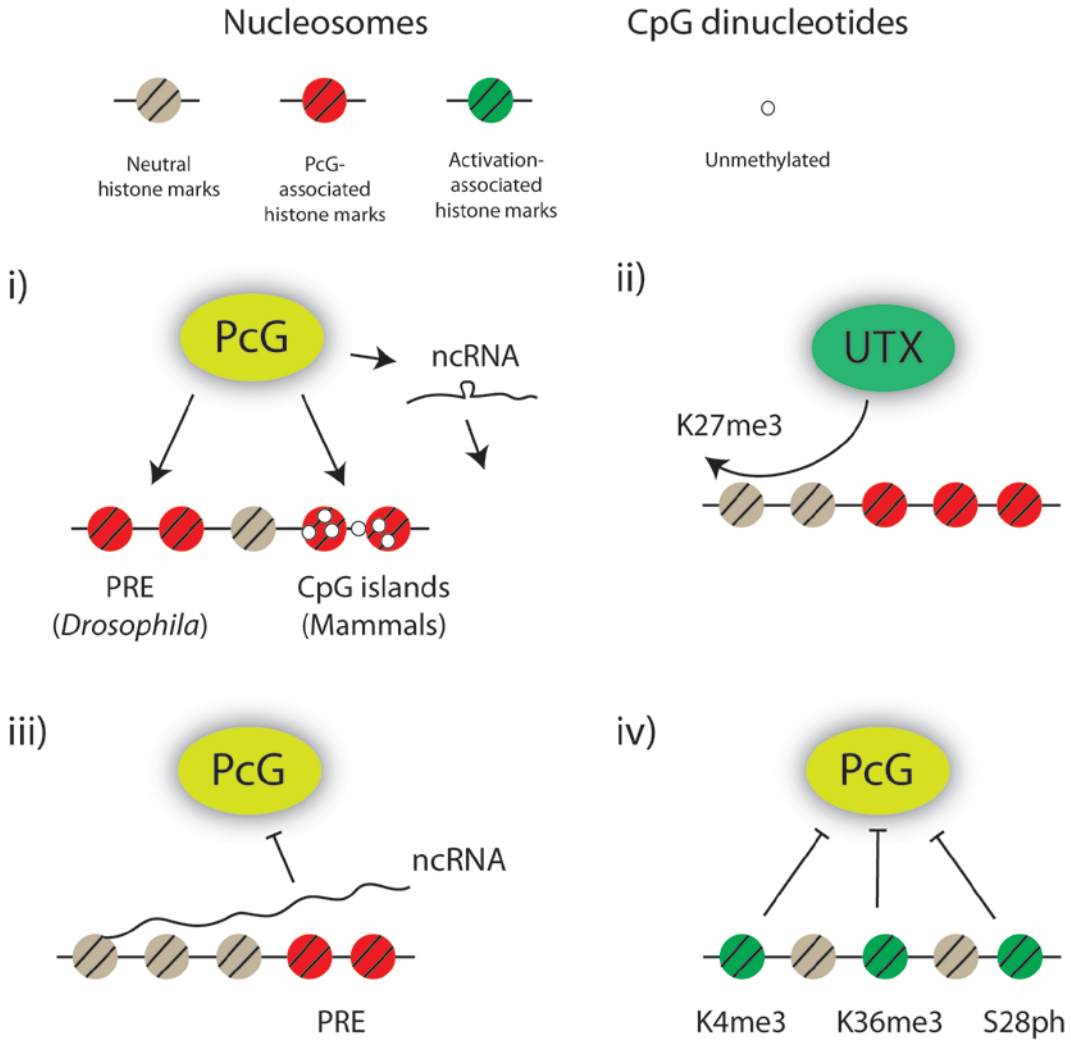


Figure 1.7 - Targeting and dynamics of PcG repression. i) Polycomb response elements (PREs) have been characterised in *Drosophila* as binding sites for PcG proteins although few have been identified in mammals where CpG islands have been linked to PcG targeting. Non-coding RNA (ncRNA) may direct PcG to certain loci. ii) PcG mediated histone marks are reversible, exemplified by a histone demethylase UTX that removes K27me3. iii) Transcription of ncRNA across a PcG binding site is thought to relieve polycomb binding and repression. iv) Certain marks on the histone H3 tail are inhibitory to PcG binding and/or function.

Transcription is thought to have a dramatic effect on PcG binding and repression. In studies of transcriptional memory of reporter genes induced by fly PREs, it was noticed that a transient burst of transcription at a PRE was enough to ablate PcG proteins and induce a memory of active transcription of the nearby reporter gene (Figure 1.7iii)(Cavalli and Paro, 1998; Cavalli and Paro, 1999). These fundamental observations were supported by further studies that suggested that transcription through PRE elements results in their loss of silencing activity in reporter constructs (Bender and Fitzgerald, 2002; Hogga and Karch,

2002; Rank *et al.,* 2002). Next, endogenous intergenic RNAs were identified that covered PREs, and their expression was shown to negatively correlate with the repression induced by the PRE, implying a role in activation (Schmitt *et al.,* 2005). Mammalian and fly *Hox* clusters have been particularly associated with this type of ncRNA, and its expression has been correlated with gene activation during spatial and temporal co-linear activation (Sessa *et al.,* 2007; Rinn *et al.,* 2007; Sasaki *et al.,* 2007). In mammals, there is abundant antisense ncRNA expression, in a pattern that resembles the neighbouring *Hox* genes on the sense strand, further suggesting a role in gene activation (Sessa *et al.,* 2007; Rinn *et al.,* 2007). The molecular function of this transcription is unclear. Specific ncRNAs, in both fly and vertebrates, have been suggested to directly bind to TrxG proteins and positively influence their recruitment in *cis* to genomic regions (Sanchez-Elsner *et al.,* 2006; Wang *et al.,* 2011).

The binding and activity of the PRC2 complex is also inhibited by certain histone modifications (Figure 1.7iv). PRC2 activity is inhibited *in vitro* on nucleosomes containing H3K4me3 or H3K36me3 (Schmitges *et al.,* 2011; Yuan *et al.,* 2011). The effect of K4me3 may be explained by reduced binding of the RpAp48/RpAp46 subunit of the PRC2 complex to histones containing K4me3 and/or by allosteric inhibition of the complex (Schmitges *et al.,* 2011). The molecular basis behind PRC2 inhibition by H3K36me3 is unclear but this effect is conserved for human and fly PRC2 (Schmitges *et al.,* 2011). The effect of H3K36me3 is intriguing, given the link between ncRNA and PcG erasure, as this is a marker of elongating transcription. Another suggested chromatin-based mechanism for PcG displacement is phosphorylation of serine 28 (H3S28Ph), the adjacent amino acid to the polycomb-deposited K27me3 mark on the histone H3 tail. Phosphorylation by Msk of this residue is reported to induce dissociation of PRC2 and gene activation if targeted to a PcG marked gene (Gehani *et al.,* 2010; Lau *et al.,* 2011). However, the importance of this mechanism in gene regulation is not yet clear as no physiological targets of H3S28Ph have been identified. In addition to these chromatin marks, DNA methylation has been linked to PcG displacement and this will be discussed in the next section.

1.3.6. Links between polycomb repression and DNA methylation

During the course of this project, intriguing links have been made between the polycomb system and DNA methylation in the literature. Generally speaking, promoters marked by K27me3 have low levels of DNA methylation and, genome-wide, K27me3 is depleted from regions with high levels of DNA methylation (Meissner *et al.,* 2008; Hawkins *et al.,* 2010). Large regions with lower than normal DNA methylation levels (partially methylated

domains - PMDs) often reflect a block of K27me3 modification in human cells (Lister *et al.*, 2009; Hawkins *et al.*, 2010). This negative correlation is also observed during chromatin dynamics when cells differentiate. For example, K27me3 domains are often broader in IMR90 fibroblasts than hES cells. Regions where this K27me3 'spreading' has occurred are also DNA hypomethylated in IMR90 cells relative to hES cells (Hawkins *et al.*, 2010). This negative correlation has also been observed in *Arabidopsis* tissues implying functional conservation (Weinhofer *et al.*, 2010). A negative correlation of these two epigenetic marks has also been observed at an imprinted locus where they are proposed to be mutually exclusive, each marking one allele (Lindroth *et al.*, 2008). These observations alone do not tell us about cause/consequence relationships between the two modifications but a few reports suggest that a causal link exists. Mutation of *Dnmt3a* results in postnatal neurogenesis defects (Wu *et al.*, 2010). Specifically, in *Dnmt3a* null neural stem cells (NSCs) transcriptional down-regulation occurs of many genes crucial for neural differentiation (Wu *et al.*, 2010). Dnmt3a was shown to bind to and methylate regions flanking CpG island promoters and within the gene body of active genes (Wu *et al.*, 2010). Loss of DNA methylation at these regions in *Dnmt3a* null cells was accompanied by increased binding of PRC2 proteins, increased K27me3 and transcriptional down-regulation. In addition, *in vitro* chromatin binding assays, performed by two independent groups, showed reduced PRC2 binding to heavily methylated DNA (Wu *et al.*, 2010; Bartke *et al.*, 2010). In a recent report, mapping of K27me3 in DNA hypomethylated mouse ES cells lacking Dnmt3a and Dnmt3b showed *de novo* targeting of this mark to previously DNA methylated CpG-rich regions (Lynch *et al.*, 2011). A reciprocal effect has been observed more indirectly by studying proteins involved in active demethylation. Depletion of Tet1 in ES cells and loss of *Tdg* in mouse fibroblasts leads to local increases in DNA methylation, decreased PRC2 binding and increased expression of many genes (Wu *et al.*, 2011; Cortazar *et al.*, 2011). These studies together suggest that dense DNA methylation may be inhibitory to PcG binding and could play a role in its targeting. This raises many questions including to what extent this mechanism is operative in other systems as well as how this effect is mediated at the molecular level. This handful of studies points towards a novel role for DNA methylation in genome regulation by modulating the targeting of PcG proteins to chromatin. As these studies were published during the course of this thesis project, they influenced the interpretation of my results and planning of experiments along the way.

1.4. Thesis aims

When I began this project, I identified the mechanisms by which DNA methylation contributes to genome regulation as one of the largest and important gaps in our knowledge. At the time it was often stated that the effects of DNA methylation were mediated by two mechanisms; recruiting methyl-CpG binding proteins and inhibiting transcription factor binding. I reasoned that the effects of DNA methylation were likely more numerous and complex than currently appreciated for the following reasons:

- Evidence suggests that MeCPs and inhibition of transcription factors each contribute to genome regulation by DNA methylation but neither can explain the severe phenotypes observed in DNA methyltransferase mutant embryos and cells. This is because loss of MeCPs results in relatively mild phenotypes and the number of transcription factor binding events affected by DNA methylation *in vivo* is very low. One interpretation of these observations is that DNA methylation affects the genome through additional mechanisms.
- An increasing variety of proteins were being identified, that were affected by DNA methylation in their binding to DNA, with diverse roles in genome regulation
- Numerous links were being drawn between DNA methylation and histone modifications suggesting extensive cross-talk that was not well understood
- Studies had suggested that other, non-canonical, roles of DNA methylation exist that are poorly understood, such as modulation of CTCF binding and regulation of lincRNAs (Fig. 1.6)

Therefore, my broad thesis aim was:

1. To investigate novel roles of DNA methylation in genome regulation.

Given the links between DNA methylation and histone modifications I became focussed on epigenetic cross-talk. In particular, the effect of experimental DNA hypomethylation on chromatin structure has not been well addressed. During the course of my project interesting links were emerging between DNA methylation and the PcG system, both in my work and in the literature. Therefore, a more specific aim that emerged during the project was:

2. To investigate the role of DNA methylation on PcG targeting and repression.

Chapter 2: Materials and Methods

2.1. Microbiological techniques

2.1.1. Transformation of bacteria

Chemically competent DH5α *Escherichia coli* (*E. coli*)(library or sub-cloning efficiency)(Invitrogen) were mixed with an appropriate amount of plasmid and incubated on ice for 10-30min. Bacteria were heat shocked by incubation at 42°C for 40sec, placed on ice for 2min, then 500µl SOC media (Invitrogen) was added followed by aseptic spreading onto L-agar plates supplemented with the appropriate antibiotic (ampicillin at 50mg/ml; kanamycin at 50mg/ml; chloramphenicol at 13mg/ml)(L-agar plates were prepared by HGU technical services). Where blue/white selection was required, 80µl of X-gal (25mg/ml in dimethylformamide) was spread on plates and allowed to dry prior to bacterial plating. Plates were inverted and incubated at 37°C overnight.

2.1.2. Growth of bacteria and isolation of plasmid DNA

To propagate plasmid DNA, single colonies of transformed bacteria were picked and used to inoculate 3ml LB (Luria-Bertani) broth supplemented with the appropriate antibiotic at the concentrations supplied above which was incubated at 37°C overnight with shaking at >225rpm (LB broth was prepared by HGU technical services). For maxi-preparations, 1.5ml of this culture was used to inoculate 150ml of LB supplemented with antibiotics and incubated in the same way for a further 16hr. Plasmid DNA was isolated from bacterial cultures using 'Nucleospin Plasmid' kit for 3ml cultures (Machery-Nagal) or 'GenElute HP Endotoxin-free Maxiprep Kit' for 150ml cultures (Sigma) according to manufacturer's instructions. Bacterial artificial chromosomes (BACs) were isolated from pelleted (1.5ml overnight culture) bacteria using the following method. Pellet was resuspended in 200ul P1 buffer (50mM glucose; 25mM Tris pH8; 10mM EDTA; 0.25mg/ml lysozyme (Sigma)) then cells were lysed by alkaline lysis by adding 400µl lysis buffer (200mM NaOH; 1% SDS w/v) and incubating on ice for 5min. Reaction was neutralised by adding 300µl acetate buffer (3M Potassium acetate; 11.5% glacial acetic acid v/v) and incubation on ice for 5min. Mixture was clarified by centrifugation at 13k*g at 4°C for 5min and the supernatant was used for phenol chloroform extraction.

2.2. DNA preparation and manipulation

2.2.1. Isolation of genomic DNA from mammalian cells

Genomic DNA was isolated from mammalian cells using either the 'DNeasy Blood and Tissues Kit' (Qiagen) according to manufacturer's instructions or the following method. Cell pellet (~10^7 cells) was resuspended in 300µl TE buffer and lysed by the addition of 300µl lysis buffer (20mM Tris pH8; 4mM EDTA; 20mM NaCl; 1% SDS w/v) and 80µg of proteinase K (proK) and incubation for >6hr at 55°C. Genomic DNA was isolated by phenol chloroform extraction. Where RNA-free genomic DNA was required, preps were treated with RNase cocktail (Ambion)(1µl RNase cocktail added with incubation at 37°C for 30min).

2.2.2. Restriction digest of genomic DNA

1.5 µg of RNA-free genomic DNA was incubated with restriction endonucleases HpaII (Roche), MspI (Roche) or HpyCHIV (NEB) according to the manufacturer's instructions for 2hrs. Digested DNA was resolved on a 1.5% agarose gel, stained using SYBR Gold DNA stain (Invitrogen) or ethidium bromide according to the manufacturer's instructions, then visualised using a FLA-5100 phosphoimager (Fuji).

2.2.3. Agarose gel electrophoresis and DNA extraction from agarose

Nucleic acids were resolved by electrophoresis through 0.7-2% agarose-TBE gels (Hi-pure low EEO agarose - Sigma) supplemented with 0.4µg/ml ethidium bromide where necessary. DNA samples were loaded by adding either 6x loading buffer (Promega) or orange G loading buffer (50% glycerol v/v; 5mM EDTA pH8; 0.3% orange G w/v). DNA size markers 1kb ladder (Invitrogen), 100bp ladder (Promega) and GeneRuler DNA ladder mix (Fermentas) were used to size fragments. DNA was extracted from agarose gel slices using 'Nucleospin Extract II' kit (Machery-Nagal) according to manufacturer's instructions.

2.2.4. Phenol-chloroform extraction and ethanol precipitation

The isolate DNA from mixtures an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Sigma) was added to at least 100µl of sample, mixed, allowed to settle then centrifuged (16k*g for 15min at room temperature). The DNA-containing aqueous phase was carefully taken and an equal volume of chloroform was added then the sample was mixed and centrifuged (5min at 16k*g at room temperature) and the aqueous phase taken to remove residual phenol. DNA was precipitated from the aqueous phase by adding 1/10 volume of 3M sodium acetate (Sigma) and 2.5 volumes of ice cold ethanol and incubating for 5min at room temperature or longer at -20°C if required. When precipitating small quantities of DNA, inert carriers like glycogen or GlycoBlue (Ambion) were added before precipitation according to manufacturer's instructions. Precipitated material was pelleted by centrifugation (16k*g for >15min at room temperature) and the pellet was washed at least once in 75% ethanol to remove salts and resuspended in distilled ultra-filtered water (MilliQ water) of TE buffer (10mM Tris pH8; 1mM EDTA).

2.2.5. DNA quantification and assessment

Nucleic acids were quantified and their purity was assessed using spectrophotometry. The 'Nanodrop ND-1000' was used to quantify both dsDNA and ssRNA samples. The A_{260} was measured and used to calculate the concentration of the sample (as an OD of 1 at 260nm = 50µg/ml of dsDNA). A_{260}/A_{230} and A_{260}/A_{280} ratios were also calculated to estimate to purity of nucleic acids.

2.2.6. pGEM-T Easy cloning

Taq polymerases often leave a dA 3' overhang on the end of PCR products. PCR products can therefore be cloned into a vector with 5' dT overhangs. PCR products were ligated with pGEM T Easy vector (Promega) according to the manufacturer's instructions but the ligation reactions were performed at 4°C overnight when maximum efficiency of ligation was required.

2.3. RNA preparation and analysis

2.3.1. Isolation of RNA from mammalian cells

Generally, the utmost care was taken when handling RNA. Clean gloves were worn, pipettes and work surfaces were rinsed with RNase Zap (Ambion), and RNase-free filtered pipette tips, eppendorf tubes and water were used throughout. RNA was isolated from mammalian cells and tissues using 'Trizol' reagent (Invitrogen) according to manufacturer's instructions. Briefly, Trizol was added to adherent cells directly onto the culture surface to minimise chance of RNA degradation. 1ml of Trizol was added per 25cm^2 of culture surface for mouse embryonic fibroblasts at >70% confluence. Mixture was incubated at room temperature for 5min to allow dissociation of protein-RNA complexes then 1/10 volume of bromochloropropane (BCP) was added and the sample was mixed vigorously by hand. After allowing the mixture to settle it was centrifuged (12k*g for 15min at 4°C) to separate the organic, inter- and aqueous phases. The RNA-containing aqueous phase was removed carefully to prevent contamination from the DNA-containing interphase. RNA was precipitated from the aqueous phase (in the presence of glycogen carrier if appropriate) by the addition of 0.5 initial volume of isopropanol for 10min at room temperature. RNA was pelleted by centrifugation (12k*g for 10 mins at 4°C) then the pellet was washed in 75% before resuspension in RNase-free water (DEPC treated water - Sigma). RNA concentration was determined by spectrophotometry using Nanodrop ND-1000 on RNA settings. RNA integrity was determined by ethidium bromide agarose gel electrophoresis looking for the integrity of the abundant ribosomal RNA fragments or by Bioanalyser RNA1000 chip (Agilent)(Bioanalyser analysis performed by HGU technical services). RNA was stored at -80°C.

2.3.2. Preparation of cDNA

Where possible, RNA was treated with DNase to remove any DNA contamination using 'TURBO DNase-free' (Ambion) according to manufacturer's instructions. Occasionally, when RNA quantities were limiting, the DNase step was not performed. In these circumstances, extra attention was paid to ensure no DNA contamination in subsequent

reactions using negative controls such as -RT (no reverse transcriptase). Between 100ng and 5µg of total RNA was used as a template for cDNA (complimentary DNA) synthesis by reverse transcription primed by either oligo-dT or random heximers (both Promega) using SuperscriptII or SuperscriptIII (Invitrogen) according to manufacturer's instructions. RNase inhibitor (RNaseIn - Promega) was added to the reaction to ensure no degradation of the RNA template in the early stages of the reaction. Negative control cDNA reactions were performed without the addition of the reverse transcriptase to control for DNA and aerosol contamination in downstream experiments using non-intron-spanning primers. cDNA was stored at -20°C.

2.3.3. RT-PCR

The presence of specific transcripts in steady-state cellular RNA was semi-quantitatively assessed by polymerase chain reaction (PCR) using cDNA as a template. Primers used for RT-PCR (Table 2.1) were designed using the 'Primer3' program (http://frodo.wi.mit.edu/primer3/) using the default settings to give a T_a (annealing temperature) of 55°C, a product size of <400bp and skipping a large intron where possible to avoid genomic DNA contamination. Primers were screened for the formation of multiple bands or extensive primer-dimers indicated by small size smears on the gel. PCR amplification was performed from cDNA in 20µl reactions as follows: 1x PCR buffer (Invitrogen), 2mM $MgCl_2$, dNTPs (0.5mM each), 150nM each primer, 0.6u *Taq* (*Thermus aquaticus*) DNA polymerase (Invitrogen) and an appropriate volume of cDNA. The amount of cDNA template used depends on the amount of RNA used in the cDNA synthesis reaction and the abundance of the template in the cDNA pool. Reactions were incubated in a thermocycler using the following program; 95°C for 5min followed by 25-35 cycles of 95°C for 20sec, 55ºC for 20sec, 72°C for 30sec. Reactions were then incubated for 72°C for 5 min to complete any polymerase elongations. PCR products were resolved by agarose gel electrophoresis and visualised using ethidium bromide staining and UV trans-illumination.

2.4. Protein preparation and analysis

2.4.1. Cell extracts

Whole cell extracts were prepared by lysing cells in Laemmli buffer (60mM Tris pH6.8; 10% glycerol; 2% SDS w/v; 0.01% bromphenol blue w/v; 10mM DTT). The lysis mixture was sonicated using an 'Soniprep 150' probe sonicator (MSE)(3 pulses of 30s at 10µm amplitude), centrifuged to remove debris (5min at 20k*g at room temperature) and then boiled at 95°C for 5min to ensure denaturation of proteins. For soluble chromatin extracts the input for N-ChIP (see 2.6.1) was used. 2x Laemmli buffer was added and extract was boiled at 95°C for 5min to ensure denaturation of proteins. Protein concentration of extracts was measured in duplicate by methanol precipitation from Laemmli buffer followed by Bradford assay. An aliquot of extract was made up to 150µl with dH_20 then 600µl methanol, 150µl chloroform and 450µl dH_20 were sequentially added and vortex mixed. The samples were centrifuged (15k*g for 2min at room temperature) and the aqueous phase was discarded. 650µl methanol was added and the sample was mixed by inversion. Precipitated protein was pelleted by centrifugation (15k*g for 2min at room temperature) and the pellet was air dried, resuspended in 50µl 1M NaOH and protein concentration was measured against BSA (bovine serum albumin) standards using Bradford reagent (BioRad) according to manufacturer's instructions. Dilution factor was taken into account when calculating protein concentration of extracts.

2.4.2. SDS-PAGE

Protein was resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Precast NuPAGE 4-12% gradient gels (Invitrogen) were loaded with samples in Laemmli buffer and pre-stained protein ladder (Fermentas) for sizing. Gels were run in a NuPAGE tank (Invitrogen) in NuPAGE SDS running buffer (Invitrogen) at ~160V.

2.4.3. Western blotting

Following electrophoresis gels were equilibrated in transfer buffer (25mM Tris; 200mM glycine; 20% methanol v/v) for 5min then transferred to polyvinylidene fluoride (PVDF) membrane (BioRad) using the GenieBlotter (Idea Scientific) according to manufacturer's instructions. Membranes were blocked with PBS (phospho-buffered saline) with 0.1% Tween20 (Sigma) and 5% non-fat dry milk proteins (Marvel)(PBST-milk) for at least 30min. Primary antibody (α-H3, Abcam, ab1791-100, 1:1000; α-H3K4me3, Upstate, 07-473, 1:1000; α-H3K27me3, Upstate, 07-449, 1:1000; α-H3K4me2, Upstate, 07-030, 1:1000; α-H4K20me3, Abcam, ab9053-100, 1:1000; α-H2AK119ub, Upstate, 05-678, 1:1000; α-Ezh2, Upstate, 17-662, 1:1000; α-Actinβ, Abcam, ab3280, 1:5000; α-Eed, Upstate, 05-1320, 1:5000; α-Suz12, Cell signalling, D39F6#3737S, 1:1000; α-H3K27ac, Abcam, ab4729, 1:1000; α-H3S28ph, Upstate, 07-145, 10µg) and secondary antibody (α-rabbit IgG-HRP (horse-radish peroxidase) conjugate, Sigma, 1:80,000; α-mouse IgG-HRP conjugate, Sigma, 1:10,000; α-mouse IgM-HRP conjugate, Sigma, 1:5000) incubations were performed in PBST-milk for at least 1hr each at room temperature, although often an overnight incubation at 4°C was used for the primary antibody. Membrane was washed extensively with PBST between incubations then antibody bound regions detected using 'SuperSignal Western Pico Reagent' chemi-luminescence reagent (Pierce) and 'ECL Hyperfilm' (Amersham).

2.5. Mammalian cell culture

2.5.1. Culture of mouse embryonic fibroblasts

Dnmt1$^{+/+}$ and *Dnmt1*$^{-/-}$ (as abbreviated for clarity) mouse embryonic fibroblasts are of the genotype *trp53*$^{-/-}$ and *trp53*$^{-/-}$*;dnmt1*$^{n/n}$ respectively and were a gift from Howard Cedar (Lande-Diner *et al.*, 2007). All mouse embryonic fibroblasts were cultured on cell culture grade tissue culture flasks (Gibco) in a humidified atmosphere at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Media (DMEM - Gibco) supplemented with 10% foetal calf serum (HGU technical services), 1x non-essential amino acids (Sigma), 1mM sodium pyruvate (Sigma), 1000U/ml penicillin and 650µg/ml streptomycin. Cells were passaged every two days or when necessary by rinsing with PBS to remove media and adding 1x trypsin-EDTA v/v (Sigma) and incubating for 2min at 37°C to allow cells to become dislodged with gentle agitation and then seeding into fresh flasks and media (~1:6 split ratio for cells used here). Passage number of cells was kept as low as possible. Aliquots of cells were frozen in liquid nitrogen in 1ml culture media supplemented with 10% dimethyl sulfoxide (DMSO) v/v after >1hr at -20°C and >5hr at -80°C. Upon recovery from liquid nitrogen storage, cells were thawed quickly by placing in a 37°C water bath then added drop wise to 10ml of culture media, pelleted by centrifugation to remove DMSO, and resuspended in fresh media and seeded. For treatment with 5-Aza-2'-deoxycytidine (5-Aza dC), 5-Aza dC (Sigma) was dissolved in DMSO at 10mM. 5-Aza dC or DMSO (control) was added to fresh culture media which was used to replace culture media of cells at 40-60% confluency to allow cells to divide in order to incorporate the drug into DNA. As 5-Aza dC is unstable in solution, media was changed daily and fresh drug added.

2.5.2. Derivation of mouse embryonic fibroblasts

Mouse embryos (CD1 E12.5) were dissected from the uterus and their deciduas, decapitated and stripped of visceral organs in ice cold PBS supplemented with pen-strep. Carcasses were minced with a scalpel and placed in 1x trypsin-EDTA (Sigma) solution at 37ºC for 30min to dissociate tissues. Solution was pipette mixed, sedimented by centrifugation at 500*g for 3min before the supernatant was plated in culture media. Tissue from multiple embryos was pooled.

2.6. Experimental procedures

2.6.1. Native chromatin immunoprecipitation (N-ChIP)

The following buffers and solutions were used during this protocol: NBA (85mM NaCl; 5.5% sucrose w/v; 10mM Tris pH7.5; 0.2mM EDTA; 0.2mM PMSF (phenylmethanesulfonylfluoride); 1mM DTT; complete EDTA-free protease inhibitor tablet (Roche)), NBB (as NBA supplemented with 0.15% NP40 (Sigma) v/v), NBR (85mM NaCl; 5.5% sucrose w/v; 10mM Tris pH7.5; 3mM $MgCl_2$; 1.5mM $CaCl_2$; 0.2mM PMSF; 1mM DTT; complete EDTA-free protease inhibitor tablet (Roche)), STOP buffer (215mM NaCl; 10mM Tris pH8; 20mM EDTA; 5.5 % sucrose w/v, 2% TritonX 100 (Sigma); 0.2mM PMSF; complete EDTA-free protease inhibitor tablet (Roche)), N-ChIP wash buffer (150mM NaCl; 10mM Tris pH8; 2mM EDTA; 1% NP40; 1% sodium deoxycholate w/v), elution buffer (0.1M $NaHCO_3$, 1% SDS), block solution (1x PBS, 0.5% BSA w/v), micrococcal nuclease buffer (50% glycerol v/v (Sigma), 10mM Tris pH7.6 and 50mM NaCl), sonication buffer (2M NaCl and 5M urea). All steps were performed on ice or at 4°C where possible unless otherwise stated. Approximately $5*10^6$ cells were harvested per IP. Cells were cultured in 25cm diameter dishes to facilitate harvesting. Cells were harvested at <90% confluency using a large cell scraper, pelleted by centrifugation (600*g for 4min at 4°C), rinsed twice in ice cold PBS and resuspended well in 5ml NBA buffer. Cell membranes were punctured by addition of 5ml NBB and incubation on ice for 4min. Nuclei were pelleted by centrifugation (2600*g for 4min at 4°C), resuspended in 10ml NBR to remove residual NBB buffer, pelleted and then resuspended in 1ml NBR (or variable depending on pellet size). Chromatin content was inferred by approximating DNA content and averaging across triplicates: 5µl of nuclei suspension was treated with DNase (RQ1, Promega) in 95µl NBR for 10min at room temp before adding 400µl sonication buffer, to disrupt DNA:chromatin interactions, followed by measuring the A_{260} by spectrophotometry against an NBR blank. According to the average measurement nuclei were distributed into 500µl aliquots in NBR each containing 400µg DNA by this measurement. For ChIP-seq, where smaller fragments are required, aliquots of 300µg DNA in 500µl were used. Nuclei aliquots were incubated at room temperature for 2min with 1µl RNase cocktail (Ambion) then digested with 25U of Mnase (micrococcal nuclease)(Worthington) for exactly 10min at room temperature. Concentration of Mnase required to produce appropriate fragment size

was titrated on each cell line used. The Mnase reaction was stopped by the addition of 500µl stop buffer and tubes were placed on ice overnight. The size range of chromatin fragments was confirmed for every replicate by performing the following. 20µl released chromatin was centrifuged (16k*g for 5min at 4°C) to remove nuclear debris and the supernatant was treated with 40µg proteinase K (proK) for 30 min at 55ºC. Sample was mixed with Orange G loading dye and run on a 1.3% agarose gel stained with ethidium bromide against a 123bp ladder (Invitrogen). Mono- di- and tri-nucleosome sized fragments should be the most abundant species. The following day, the released chromatin was centrifuged (16k*g for 5min at 4°C) and the supernatant was collected and pooled for each cell line or condition used. 100µl of each was set aside as input and either stored at -20ºC or one ice for the rest of the day. Protein-A Dynabeads (Invitrogen) were washed twice in block solution. Chromatin was pre-cleared by incubating with 15µl Dynabeads for 2 hr at 4ºC on rotating wheel. In parallel, antibodies (rabbit IgG, Santa Cruz, sc-2027, 10µg; α-H3K4me3, Upstate, 07-473, 5µg; α-H3K27me3, Upstate, 07-449, 10µg; α-H3K27ac, Abcam, ab4729, 10µg; α-H3S28ph, Upstate, 07-145, 10µg) were bound to beads by incubating 30µl Dynabeads in 200 µl block solution with antibody on rotating wheel at 4ºC for 2 hr. Antibody bound beads were washed once with block solution before addition of pre-cleared chromatin. Immunoprecipitation was performed for 3 hr on rotating wheel at 4ºC. Beads were washed three times in N-ChIP wash buffer for 10min at 4ºC on rotating wheel and then rinsed in TE. DNA was eluted in one of two ways. 100µl elution buffer warmed to 37ºC was added to beads and they were rotated for 15 min at room temp. Elution was repeated using 50µl elution buffer and the two elutions were pooled. DNA was isolated from eluate and inputs by treatment with 40µg proK for 1hr at 55ºC and purification using PCR clean up kit (Qiagen) according to manufacturer's instructions. Alternatively, when maximum DNA yields were required, beads were resuspended in 200µl ProK digestion buffer (50mM Tris pH8; 10mM EDTA; 0.5% SDS w/v), 40µg ProK was added to bead suspension and to inputs, and reactions were incubated at 56°C overnight with intermittent agitation, then DNA was isolated using PCR clean up kit (Qiagen). DNA was eluted from columns in 50µl dH_20 and quantified by Nanodrop ND-1000 spectrophotometry. Controls for N-ChIP experiments included a species matched IgG to control for background, negative and positive genomic regions to show specificity, and the use of input DNA to control for technical variations between cell lines or treatments. Antibodies were tested by western blot for recognition of histone-sized proteins only and by specific enrichment at positive control regions by N-ChIP. Enrichments were measured by qPCR (see 2.6.4) and/or microarray (see 2.7 and 2.8).

2.6.2. Cross-linked chromatin immunoprecipitation (X-ChIP)

The following buffers and solutions were used during this protocol: Cytoplasmic lysis buffer (5mM PIPES ph8; 85mM KCl; 0.5% NP40 v/v (Sigma); complete EDTA-free protease inhibitor tablet (Roche)), PBS-RIPA (1xPBS; 1% NP40 v/v; 0.5% sodium deoxycholate w/v; 0.1% SDS w/v; complete EDTA-free protease inhibitor tablet (Roche)), X-ChIP wash buffer (150mM NaCl; 10mM Tris pH8; 2mM EDTA; 1% NP40; 0.1% sodium deoxycholate w/v), LiCl wash buffer (100mM Tris pH7.5; 500mM LiCl; 1% NP40; 1% sodium deoxycholate), block solution (1x PBS, 0.5% BSA w/v), sonication buffer (2M NaCl and 5M urea). All steps were performed on ice or at 4°C where possible unless otherwise stated. Approximately $5*10^6$ cells were harvested per IP. Cells were cultured in 25cm diameter dishes to facilitate fixation and harvesting. Formaldehyde (36.5% from Sigma) was added to 1% final concentration directly to culture media and dishes were incubated at room temperature with mild agitation for exactly 10min. Formaldehyde was quenched by the addition of glycine to 125mM final concentration and dished were incubated for 3min at room temperature. Cells were rinsed twice in ice cold PBS before being dislodged with a large cell scraper in ice cold PBS with added protease inhibitors (complete EDTA-free protease inhibitor tablet (Roche)). Cells were pelleted by centrifugation (600*g for 4min at 4°C) and fully resuspended in 5ml cytoplasmic lysis buffer and left on ice for 10min. Nuclei were pelleted by centrifugation (2600*g for 4min at 4°C) and fully resuspended in 600µl PBS-RIPA and incubated on ice for at least 20min. Extract was sonicated using a 'Soniprep 150' probe sonicator (MSE) at 6µm amplitude for 30sec on/off cycles for 5 cycles while on ice between sonications (this was determined by testing a range of sonication conditions). Extract was centrifuged (8k*g for 3min at 4°C) to remove nuclear debris and the supernatants from the same cell lines or treatments were combined. The size range of chromatin fragments was confirmed for every replicate by performing the following. 15µl extract was treated with 1µl RNase cocktail (Ambion) at 37°C for 15min then with 40µg proteinase K (proK) for >1hr at 65ºC to reverse cross-links. Sample was centrifuged (16k*g for 3min at room temperature) then the supernatant was mixed with Orange G loading dye and resolved on a 1.3% agarose gel stained with ethidium bromide against DNA ladders. The DNA smear should be between 200 and 1000bp. Chromatin content was inferred by approximating DNA content and averaging across triplicates: 5µl of nuclei suspension was added to 495µl sonication buffer, to disrupt DNA:chromatin interactions, followed by measuring the A_{260} by spectrophotometry against an sonication buffer blank. According to the average measurement extract was distributed into 1ml aliquots in PBS-RIPA each

containing 300µg DNA by this measurement. 100ul of extract for each cell line or treatment was put aside as input and stored on ice. Antibody (mouse IgG, Santa Cruz, sc-2025, 5µg; α-Ezh2, Upstate, 17-662, 5µg) was added to each tube and they were incubated on a rotating platform at 4°C overnight. 35µl sheep α-mouse IgG conjugated Dynabeads (M280 - Invitrogen) were washed twice in block solution and added to antibody-extract solution and tubes were rotated at 4°C for 2hr. Beads were washed for 10min rotating at 4°C once in PBS-RIPA, once in X-ChIP wash buffer and once in LiCl wash buffer before being rinsed in TE. Beads were resuspended in 200µl ProK digestion buffer (50mM Tris pH8; 10mM EDTA; 0.5% SDS w/v) and treated, together with inputs, with 1µl RNase cocktail (Ambion) for 20min at 37°C. 40µg ProK was added to bead suspension and to inputs, and reactions were incubated at 65°C for the rest of the day, then NaCl was added to 200mM and tubes were incubated overnight at 65°C with intermittent agitation to reverse cross-links. DNA was isolated using PCR clean up kit (Qiagen) and eluted from columns in 50µl dH_20 and quantified by Nanodrop ND-1000 spectrophotometry. Controls for X-ChIP experiments included a species matched IgG to control for background, negative and positive genomic regions to show specificity, and the use of input DNA to control for technical variations between cell lines or treatments. Antibodies were tested by western blot for recognition of proteins of the expected size and by specific enrichment at positive control regions by X-ChIP. Enrichments were measured by qPCR (see 2.6.4).

2.6.3. Methylated DNA immunoprecipitation (MeDIP)

The following buffers are used during this protocol: Block solution (1x PBS, 0.5% BSA w/v), 10x IP buffer (100mM sodium phosphate pH7; 1.4M NaCl; 0.5% TritonX-100 v/v), ProK digestion buffer (50mM Tris pH8; 10mM EDTA; 0.5% SDS w/v). 20µg of genomic DNA (see 2.2.1) was diluted in 400µl TE and sonicated in 1.5ml eppendorf tubes using a 'Bioruptor' water bath sonicator (Diagenode) for 6 cycles of 30sec on/off on the 'high' setting with a shallow volume of ice in the water bath to prevent overheating. Ice was replaced regularly. 5µl was loaded on a 2% agarose gel containing ethidium bromide to check that the fragment size of each replicate was between 300 and 1000bp after sonication. 400ng of sonicated DNA is put aside as input. 4µg DNA is diluted in 450µl TE for each immunoprecipitation (IP) to be performed. Sonicated DNA (including input) was denatured by heating to 100°C for 10min then quick chilling on ice, to facilitate recognition by 5-mC antibody. 51µl of 10x IP buffer was added along with 10µg of antibody (mouse IgG, Santa Cruz, sc-2025; α-5mC, Eurogentec, #BI-MECY-1000) and tubes were rotated for 3hr at 4°C.

40µl of sheep α-mouse IgG-conjugated Dynabeads (M280, Invitrogen) per IP were washed twice in block solution then added to DNA-antibody mixture and tube were rotated for a further 2hr at 4°C. Beads were washed 3 times with 1ml 1x IP buffer each for 10min at room temperature. Beads were resuspended in 200µl ProK digestion buffer, 60µg ProK was added and tubes were incubated for >2hr at 56°C with intermittent agitation. DNA from IPs and inputs was isolated using PCR clean up kit (Qiagen) and eluted in TE or dH_20. Controls used for MeDIP include a species matched IgG to control for background, positive and negative control regions, and the use of input DNA that has been treated the same way as IP DNA. Enrichments were measured by qPCR (see 2.6.4) and/or microarray (see 2.7 and 2.8).

2.6.4. Quantitative polymerase chain reaction (qPCR)

DNA and cDNA was quantified using quantitative real-time polymerase chain reaction (qPCR). Reactions were performed in duplicate in white 96-well plates (Eurogentec) in either a CFX96 (BioRad) or LightCycler (Roche) and were 20µl in total composed of 10µl '2x Brilliant II qPCR master mix' (Agilent), 150nM each primer, DNA or cDNA template in MilliQ dH_20. Reactions were performed using the following program: 95°C for 5min, then cycles of 94°C for 15sec, 55°C for 20sec and 72°C for 20sec, followed by a melt curve ramping from 60-95°C. Fluorescence signal of the SYBR green dye was measured after each cycle of PCR. Primers for qPCR were designed using the 'Primer3' program (http://frodo.wi.mit.edu/primer3/) to produce primers with an annealing temperature of 55°C and a product size of <150bp. For cDNA, primers were designed to span a large intron where possible to avoid genomic DNA contamination. All primers used are shown in table 2.1. All primers were screened for poor efficiency and the formation of primer dimers or secondary bands by serial dilution and melt-curve analyses. For analysis of cDNA, a standard curve was generated by 10-fold serial dilution of a cDNA sample from a tissue where the gene is suitably abundant. This is also performed for a housekeeping gene, such as *actb*, to normalise for loading and overall efficiency of the cDNA synthesis reaction. When normalising to a loading control gene one must take into account its error in the calculation of relative quantities. This was done by propagation of errors as described at "http://ion.chem.usu.edu/~sbialkow/Classes/3600/Overheads/Propagation/Prop.html". This method allows for the relative quantity of cDNA species to be calculated between cell lines or treatments. When analysing DNA in immunoprecipitation experiments, such as N-ChIP, X-ChIP and MeDIP, relative quantitation was also performed. A standard curve was produced using 10-fold serial dilutions of the input DNA which itself is known to be 10% of

the material used for each IP. In this way the amount of DNA template in each IP sample can be expressed relative to the amount in the input, controlling for overall variations in amount of material used in different treatments. Where technical or biological replicates were performed, the qPCR error (error between the two qPCR replicates) was ignored and the standard error was calculated between the technical or biological replicates. Where only one technical or biological replicate was performed the error is calculated as the qPCR error. Error bars denote the standard error of the mean (S.E.M) for each sample, propagated if necessary.

Primer pair name	Application	Target	Sequence (5'-3')
RT_Hoxc4	cDNA	-	CTACCCTGAGCGTCAGTATAGC CGCAGAGCGACTGTGATTTCT
RT_Hoxc5	cDNA	-	CATGAGCCACGAGACGGATG CGAGTGAGGTAGCGGTTAAAATG
RT_Hoxc6	cDNA	-	AATTCCACCGCCTATGATCCA ACATTCTCCTGTGGCGAATAAAA
RT_Hoxc8	cDNA	-	CTTCGTCAACCCCCTGTTTTC GTCTTGGACGTGGTGCGAG
RT_Hoxc9	cDNA	-	ACTCGCTCATCTCTCACGACA AGGACGGAAAATCGCTACAGT
RT_Hoxc10	cDNA	-	ATGACATGCCCTCGCAATGTA CCCCGCAGTTGAAGTCACTC
RT_Hoxc11	cDNA	-	GAGAACACGAATCCCAGCTC CGGATCTGGAATTTCGAATAAG
RT_Hoxc12	cDNA	-	ATGGGCGAGCATAATCTCCTG CGTGGGTAGGACAGCGAAG
RT_Hoxc13	cDNA	-	TACCAGCACTGGGCTCTTTC CTCACTTCGGGCTGTAGAGG
RT_Hoxa11	cDNA	-	TTTGATGAGCGTGGTCCCTG AGGAGTAGGAGTATGTCATTGGG
RT_Hoxa13	cDNA	-	CTGCCCTACGGCTACTTCG CGGTGTCCATGTACTTGTCG
RT_Hoxa10	cDNA	-	TAGAGGTTGAGATCTTTCTCAG TCCCTGATTAAACACAGCCCAG
RT_Hoxb9	cDNA	-	AGGAAGCGAGGACAAAGAGAG CTTGTCTCTCACTCAGATTGA
RT_Actin	cDNA	Actb	AGAGCTATGAGCTGCCTGACG TGTGTTGGCATAGAGGTCTTTACG
RT_Vax2	cDNA	-	AGTATGTGGTGGGCCGAGAG CTGGTCCTTCTTCTGCTTGG
RT_Foxq1	cDNA	-	TTCGGAAAAGCGTCTCTCTC GAGGTCGTAGGAAGCGAGTG
RT_Wnt10a	cDNA	-	CCTCACAGAGACATCCATGC CCGTGGCATTTGCACTTAC
RT_Hand2	cDNA	-	TCAAGGCGGAGATCAAGAAG TGGTTTTCTTGTCGTTGCTG
RT_Tex19	cDNA	Tex19.1	AAAATGGGCCACCCACATCTC CCACTGGCCCTTGGACCAGAC
RT_Gapdh	cDNA	-	ACCCAGAAGACTGTGGATGG GGTCCTCAGTGTAGCCCAAG
ChIP_Hoxc4	DNA	Promoter	CATCGATCCGAAATTTCCTC TACAGCTCCTGGTGGTGATG
ChIP_Hoxc5	DNA	Promoter	CGGCTTCCATCACTAACCTC ATGCGCTCTTCCCAAATAAG
ChIP_Hoxc8	DNA	Promoter	CCTATTACGACTGCCGGTTC CGTGGTGGAAGAAGTCTTGG
ChIP_Hoxc9	DNA	Promoter	TCAGTCTGGGCTCCAAAGTC AGAGGTAGCCTCCCCAGAAC
ChIP_Hoxc10	DNA	Promoter	GCTAGGTGGCGCTGTTACTC CCAATGGGATTTGAAAATGG
ChIP_Hoxc12	DNA	Promoter	AAATCTGTTCCTGCCTTTCG TCTACTCACGTGGGCGCTAC
ChIP_Hoxc13	DNA	Promoter	AGACCCAGGCTTAGCATCAC TAAGAAATCCGGCGACTCC
ChIP_Actin	DNA	Actb promoter	GAATGTGGCTGCAAAGAGTCTAC CTTCGCTCTCTCGTGGCTAGTA
ChIP_Cphx	DNA	Promoter	AGATTGGGTGGGGTGCTTAG CCTCCAATCTCAGACCTTGC
ChIP_Chdh	DNA	Promoter	TCAGCTAAACGGATGGGAAC

			GGGCCATGTTGCTAGTGTTG
ChIP_Sytl3	DNA	Promoter	CTGGGGAAATTTTGCTACCC TAAATGATGCATGGCTGCTG
ChIP_Vax2	DNA	Promoter	CCTCCCCTTGTCTCTCTTCC CTGGGGTTCCAGGACTCTG
ChIP_Tex19	DNA	Promoter	CGACTTTTCCAAACAAGATGTG AGTCACTCTGGCTCTGAAGG
ChIP_Tbr1	DNA	Promoter	AGCTAACCAATGGCAAGAGC TTCCCTGGGAGAAGAGATTG
ChIP_Gata6	DNA	Promoter	CCTTCCCATACACCACAACC CCCCTCCTTCCAAATTAAGC
ChIP_Zic5	DNA	Promoter	AGCCTGAACCAGAGCCAATC ACATTTGAGCTGCACAGTGG
ChIP_Hand2	DNA	Promoter	CTCGGCAATTAGCAACGTG CGCTCGGGTTAATATATGTCG
ChIP_Wnt10a	DNA	Promoter	TTCAACCAGGAGGGTGAGAG GGGTGCTTTGAGACATGACC
ChIP_Pcdh18	DNA	Intragenic CpG island (Illingworth)	CAGACCGAGGCTCACCTG ATCGGGTAGAGCACGAAGG
ChIP_Pcdhb20	DNA	Intragenic CpG island (Illingworth)	GAACCTGGCTACCTGGTCAC CCTGGCTCTGTAGCCTTGAG
ChIP_Dazl	DNA	Promoter	CCACCTTCGAGGTTTTACCA TCTCCCCACTTCTCTGTGCT
ChIP_Gapdh_prom	DNA	Promoter	TCCCTAGACCCGTACAGTGC CTCTGCTCCTCCCTGTTCC
ChIP_Gapdh_exon	DNA	Exon	CAGCTCTCCCCATACATACA CATTTCTTATCTTACCCTGCC
ChIP_Tex19.1_exon	DNA	Exon	TGACCCTTACTCCCTGATGC GGCAACCTCTTGCTCTTCAC
ChIP_Hoxc11-c10_inter	DNA	Intergenic between Hoxc11 and Hoxc10	GCCGAGTTCAAAGTTTCCTG CCACGTGCGCTATAAACTTG
ChIP_HoxA_CBS5	DNA	CTCF binding site (intergenic)	GTCCCTTGGAGGCTTTCAG AGCCTTTCTGGAATGGGATT
ChIP_Hoxc8-c9_CTCF	DNA	CTCF binding site (intergenic)	GCGCTGTTGCTCAATGTTAG CAGGGCCTTTCTGACTGAAG
cMyc - 5kb	DNA	5kb upstream of c-Myc gene	CAGGCAATGTAGGTTCTTGG CGATGTGATCCCTTCCCATA
cMyc_5'	DNA	Promoter	CAAGGAAGCATCTTCCCAGA TACTGGCCACAGAGACCACA
Pax6_6a	DNA	50kb upstream of Pax6	CCAGTGCTCTGGGCTACAAT AAGACGCCAGGAAGAGGATT
BS_Oct4_prom	BS	Promoter	GTTTTGGATATGGGTTGAAA CCCACCTAATAAAAATAAAAAAAC
BS_Oct4_prom (nested)	BS	Promoter	TGAGGAGTGGTTTTAGAAATAA TAAAACCAAATATCCAACCA
BS_Tex19_prox	BS	Proximal promoter	GGAAGTATATAGGGTATTTA AAAATATTAAAACTCTAACCCCC
BS_Tex19_prox (nested)	BS	Proximal promoter	GTTTTTTGTTGTTTTGGGAT CTATTCACTCTAAACAAACT
BS_Tex19_distal	BS	Distal promoter	GGTTTTGTTTTTTGTTGTTG CATTTACATATCTCCCATAAAATC
BS_Tex19_distal (nested)	BS	Distal promoter	TTATTAAAGAGATAGGGAAGAAG ATCCCAAAACAACAAAAAAC
BS_Pcdhb18_CGI	BS	Intragenic CpG island (Illingworth)	TTATATTATTTATTYGTTGTTGTTA AACAACCAAACATTCTATCCA
BS_Pcdhb18_CGI (nested)	BS	Intragenic CpG island (Illingworth)	GTTAATTTAAGATTTATAGATG AACCAAACATTCTATCCAAA
BS_Hoxc4	BS	Promoter	TATATGGTGAAAGTAATTTTATAGGGT AAAATAATCAAAAACTAACTAACTACAAAC
BS_Hoxc4 (nested)	BS	Promoter	ATTTTAGAAATTAATGATTATGAGTT TCAAAAACTAACTAACTACAAACTAAC
BS_Hoxc9	BS	Promoter	GTTTGTTTGTTTGTTTGAGATTTT CCCAAAACCAAAATTACCCTATACT
BS_Hoxc9 (nested)	BS	Promoter	GATAATTTAGGTTGGGAGGGAGTAT CCCTATACTCTAATTCAATAACTCTAAAC
BS_Hoxc10_intra	BS	CpG island 1st exon	TTAAGGAGGAGAATGTTTGTTGTATG TTCAAAACCTTTAAATTCCTTAAACC
BS_Hoxc10_intra	BS	CpG island 1st exon	TTGTTGTATGTATAGTGTAGAGAAG TTAAACCTCTACTCAATTCTCC

(nested)			
BS_Hoxc11	BS	Promoter	TTTTTTTYGTTTTGTGGAATGGAG CAACTAAACAAATAAAAATTAAAAATAC
BS_Hoxc11 (nested)	BS	Promoter	TTTTGTGGAATGGAGAGGAGG TAAAAATTAAAAATACAACTCCCTC
BS_Hoxc13	BS	Promoter	GTTTTTAAAAAGTTGGAGTAG TAACTACCCCCAAATAAATAAC
BS_Hoxc13 (nested)	BS	Promoter	AAGTTGGAGTAGATTATGTTATGA CAAATAAATAACCATAACCC
BS_Hoxa10	BS	Promoter	GTGTAAATTAGAGTGGGTGGT CAAAACATCAAAAAAAAAAAACT
BS_Hoxa10 (nested)	BS	Promoter	GTGGAGTTTAGGATTYGTTTTTTT CAAAAAAAAAAACTCCTACTACC
BS_Hoxc4-up	BS	CpG island 2kb upstream of promoter	TTAAGTGTTAATTAGGTTGTGTTG CCAAATCTCCTTATTATACT
BS_Hoxc4-up (nested)	BS	CpG island 2kb upstream of promoter	TAATTAGGTTGTGTTGGGAG TCTCCTTATTATACTTAATAAC
BS_Hoxa9-a10	BS	Intergenic	AAGTGTTTTTTGTTTTYGATGTGG TTACATCTTCCATCTTTCTATC
BS_Hoxa9-a10 (nested)	BS	Intergenic	GGGTGTTTTTTTATYGGATTTTT CATCTTTCTATCCAAATCTA
BS_Hoxa11-a13	BS	Intergenic	GAGAAAATGTTTTTGTTAAGT AATAAAAACCCRAAATTTACCC
BS_Hoxa11-a13 (nested)	BS	Intergenic	AGGAATTAAGGTTAGGAGAGA CATAAAAACRACAAACAAAAAAAC
BS_Hoxc13-c12	BS	Intergenic	AAAGATGGGAAGAAGTTAAGATT CTACAAAATATCTCTCAACT
BS_Hoxc13-c12 (nested)	BS	Intergenic	GGGAAGAAGTTAAGATTTTTTTT CTACAAATAACATATCATACC
BS_Hoxc12-c11	BS	Intergenic	GTATTTTGGAGTTGTTAGAY TTTCTTCCCTAAAATAAACC
BS_Hoxc12-c11 (nested)	BS	Intergenic	GAGTTGTTAGATTTTTGTGT CAAAACCTTCTCCCTATCTC
BS_Hoxc10-up	BS	Promoter	GGAGTTATAGGTTGATTTTTTAGT AAAAACTAACATCTTTTTTCCCC
BS_Hoxc10-up (nested)	BS	Promoter	TATAGGTTGATTTTTTAGTAAATAAG ACTAACATCTTTTTTCCCCCATC
BS_Hoxc10-c9	BS	Intergenic	AAAGAGGTAGAGAGTAAATTGTAA AAAAAAACCAAAACCACCAA
BS_Hoxc10-c9 (nested)	BS	Intergenic	AGAGTAAATTGTAATTTTGA CAAAACCACCAATTTACCCC

Table 2.1 - Primer sequences. cDNA: Primers used for expression analysis from cDNA. DNA: Primers used for DNA analysis including MeDIP, N-ChIP and X-ChIP. BS: Primers used for bisulfite sequencing. For bisulfite sequencing some primers contain mixed bases. Y: Pyrimidine (C or T). R: Purine (G or A).

2.6.5. Bisulfite sequencing

500ng genomic DNA was bisulfite treated using 'EZ Methylation Gold' kit (Zymo research) according to the manufacturer's instructions. Bisulfite primers were designed using a combination of the 'Bisearch' algorithm (http://bisearch.enzim.hu/) and by eye approaches and primers were designed to have a C or G base at their 3' end where possible to facilitate priming (Table 2.1). PCR amplification was performed in 20µl reactions as follows: 1x PCR buffer (Invitrogen), 2mM $MgCl_2$, 0.5mM each dNTPs, 150nM each primer, 0.6 units *Taq*

polymerase (Invitrogen) and 1-2µl bisulfite treated DNA. Reactions were incubated in a thermocycler at 94°C for 1 min followed by 35 cycles of 94°C for 10 sec, 50ºC for 35 sec and 72°C for 30 sec. Reactions were then incubated for 72°C for 5 min. A negative control reaction containing dH_20 in place of bisulfite-converted DNA was used for each set of reactions. PCR clean up of the first reaction was performed using 'Nucleospin Extract II' kit (Zymo research) according to manufacturer's instructions and DNA was eluted in 15 µl water. Nested PCR was performed as above using up to 10µl of eluted DNA from the first reaction and the products were resolved on a 1.5% agarose gel. Gel extraction, cloning into pGEM T Easy vector, transformation into bacteria and blue/white screening was performed as described in sections 2.1 and 2.2. Up to 24 individual white colonies were picked per reaction and grown overnight in 1ml LB broth supplemented with 50µg/ml ampicillin in 96-well blocks. Plasmid DNA was isolated by miniprep and sequenced using SP6 sequencing primer by robot (performed by HGU technical services). Each sequence was imported into the 'BioEdit' program (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and aligned to a reference sequence for the region, created by *in silico* bisulfite treatment, using the 'ClustalW' inbuilt program. The methylation status of each CpG dinucleotides from each bacterial clone was assessed by presence or absence of C to T conversion in te PCR product and visualised using 'QUMA' online program (Kumaki *et al.*, 2008)(http://quma.cdb.riken.jp/). Quality control steps were performed, such as checking that >95% of C, in the non-CpG context, had been converted to T. Care was taken to exclude any clonality in the sequences by looking for sequence variances in each clone, especially if their methylation pattern is identical. This is possible as the C to T conversion rate of the bisulfite reaction is not 100% leaving remaining Cs at different locations. Where an ambiguous base was called, the chromatogram was studied and the sequence edited if an obvious signal had been missed. The overall percentage CpG methylation was calculated for each region.

2.6.6. Immunofluorescence of fixed cells

Mouse embryonic fibroblasts were seeded on gelatinised glass coverslips at a density of $2*10^5$ cells per well of a 6-well plate and cultured for 24hr. Cells were washed twice in PBS then fixed in 4% formaldehyde (v/v in PBS)(Sigma) for 10min at room temperature with gentle rocking. Cells were rinsed 3 times in PBS to remove formaldehyde and then permealised by adding 0.1% TritonX-100 (v/v in PBS)(Sigma) for 5min at room temperature followed by 2 rinses in PBS. Coverslips were transferred to a humidified chamber and 100µl

of block solution (PBS; 0.5% BSA w/v; 0.1% Tween20 v/v (Sigma); 10% goat serum v/v) was added to cover the culture surface of the coverslips and then coverslips were incubated for 30min at room temperature. Goat serum was added to the block solution as the secondary antibodies used were raised in goat, therefore minimising background signal produced by secondary antibody binding. Primary antibodies (α-H3K27me3, Upstate, 07-449; α-H3K4me3, Abcam, ab1012; α-Ezh2, Upstate, 17-662; α-Bmi1, Upstate, 05-637) were diluted 1:100 in block solution and 100µl was added to coverslips. Thin, clean plastic strips were used to cover the coverslips to prevent them drying and coverslips were incubated overnight at 4°C. Coverslips were washed 3 times for 5min in PBS with 0.1% Tween20 v/v (Sigma). Secondary antibodies (Goat α-rabbit IgG-Alexafluor Red conjugate (Invitrogen); Goat α-mouse IgG-Alexafluor Green conjugate (Invitrogen)) were diluted 1:2000 in block solution with 2µg/ml DAPI (4',6-diamidino-2-phenylindole) and added to coverslips in the dark for 1hr at room temperature. Coverslips were washed 3 times for 5min in PBS with 0.1% Tween20 v/v (Sigma) protected from light. Coverslips were mounted in 'Vectashield' mounting media (Vector Laboratories) with added 250ng/ml DAPI and viewed using 2 or 3 colour fluorescence microscopy using defined exposure settings. Fluorescence images were taken using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics), Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives and Chroma #83000 triple band pass filter set (Chroma Technology). Images were captured using defined exposure settings to be directly comparable and fluorescence signal was quantified using 'IP Lab' software and normalised to DAPI signal.

2.6.7. Fluorescence in situ hybridisation (FISH) and interprobe distance measurement

FISH was performed with a lot of help from Shelagh Boyle. BAC clones that span the mouse *HoxA* cluster were obtained as bacterial stab cultures from 'BACPAC', which were cultured and the plasmid DNA was isolated as per section 2.1 (Clones IDs: WI1-2560F9 and WI1-2081I16). The BACs were labeled with biotin-16-dUTP (Roche) or digoxigenen-11-dUTP (Roche) by nick translation as follows: A 20µl reaction was set-up containing 50mM Tris pH7.5, 10mM $MgSO_4$, 0.1mM DTT, 50µg/ml BSA fraction V (Sigma), 62.5µM each dATP, dCTP and dGTP, either 62.5µM bio-16-dUTP or 37.5µM digoxigenen-11-dUTP and 25µM dTTP, 500-1000ng BAC, 1µl of 1:500 dilution of DNaseI (Invitrogen) and 1µl DNA polymeraseI (Invitrogen). Reaction was incubated at 16°C for 90min then stopped by adding

SDS to 2% w/v and EDTA to 60µM and the DNA was purified using a 'QuickSpin' column (Pharmacia). Labeling of BACs was checked using α-dig-alkaline phosphatase (AP) conjugated antibody and streptavidin-AP and BCIP/NBT AP substrate kit (Vector laboratories) against standards. Cell nuclei were prepared by cytoplasmic lysis in hypotonic solution (70mM KCl; 9.5mM trisodium citrate) for 10min at room temperature then fixed in 3:1 methanol:acetic acid for 1hr at room temperature. Fixed nuclei were dropped onto glass slides and aged for at least 24hr at room temperature. 20x SSC (saline-sodium citrate) buffer was prepared (3M NaCl; 300mM trisodium citrate; pH7 with HCl). Per slide, 100ng each labeled BAC and 5ug sonicated salmon sperm DNA (Sigma) were consolidated and dried in a SpeedVac then resuspended in 15µl hybridisation mix (50% deionised formamide v/v (Sigma); 10% 20x SSC v/v; 10% dextran sulfate v/v (Millipore); 1% Tween20 (Sigma)). Slides were treated with 100µg/ml RNaseA in 2x SSC buffer at 37°C for 1hr then washed and dehydrated through 70%, 90% and 100% EtOH for 2min each. Slides were warmed in a 70°C oven for 5min, denatured in 70% formamide in 2x SSC buffer pH7.5 for 1-1.5min then transfer to ice cold 70% EtOH and dehydrated as before. Probe and competitor DNA was denatured at 70°C for 5min then incubated at 37°C for 15min. 10µl was pipetted onto coverslips which were added to pre-warmed slides then sealed using TipTop Rubber solution (Rema) and incubated in a humidified chamber at 37°C overnight. Slides were washed 4 times for 3min in 2x SSC at 45°C then 4 times in 0.1x SSC at 60°C. Slides were briefly blocked in 40µl blocking buffer (5% non-fat dried milk powder w/v (Marvel) in 4x SSC buffer) for 5min at room temperature then each antibody incubation was performed in 40µl block solution at 37°C in a humidified chamber. Between antibody incubations slides were washed 3 times for 2min in 4x SSC with 0.1% Tween20 (Sigma) at 37°C. After the final wash coverslips were mounted on slides using 25µl 'Vectashield' (Vector Laboratories) with 250ng/ml DAPI. Antibody incubations were performed in the following order with washes between them: Sheep α-digoxigenin-FITC (1:20)(Roche); both α-sheep IgG-FITC (1:100)(Vector Laboratories) and Avidin-TxRd (1:500) (Vector Laboratories); α-avidin-biotin (1:100)(Vector Laboratories); Avidin-TxRd (1:500). Three colour fluorescence images were taken using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics), Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives and Chroma #83000 triple band pass filter set (Chroma Technology). Interprobe distance was measured in pixels using custom scripts written for 'IP Lab' software (written by Paul Perry), converted to µm and normalised to nuclear area as measured by DAPI signal. Final values were expressed as d(interprobe distance in µm)^2/r(nuclear radius in µm)^2.

2.7. Nimblegen arrays

Analyses by microarray for cDNA, N-ChIP and MeDIP were performed using Nimblegen arrays in house and by outsourcing to a service centre (Microarray Core Facility, VU University Medical Centre, Amsterdam). Data analysis was performed in house (see section 2.8).

2.7.1. Whole genome amplification

Whole genome amplification of immunoprecipitated and input DNA was performed using 'WGA2' kit (Sigma) with the following adaptations. The initial 'fragmentation' step (fragmentation buffer was added but the critical heat step was not performed) was skipped as DNA was already fragmented by either Mnase digestion or sonication. 10ng of both input and immunoprecipitated DNA were used as template for amplification. Following genome amplification, qPCR analysis was performed on a handful of control loci where the enrichments were known and compared to pre-amplification DNA to check for amplification biases. No major differences were observed (data not shown).

2.7.2. Labelling, hybridisation and scanning

All sample preparations and quality control steps were performed as per Nimblegen guides for single-colour cDNA arrays and two-colour arrays. For MeDIP, the Nimblegen ChIP-chip guide was followed. All experiments were performed in triplicate. For immunoprecipitation techniques, input DNA was arrayed along with enriched DNA by labelling them with two independent dyes, Cy3 and Cy5. Input DNA was used as appose to IgG or any other 'mock' IP as these techniques rely on very small quantities of DNA which may result in more noise, although the choice between input and mock is contentious. Briefly, for cDNA array, double-stranded cDNA was synthesised using oligo-dT primers (using 'Double-stranded cDNA synthesis kit' - Invitrogen) according to manufacturer's instructions. Double-stranded cDNA was run on a Bioanalyser RNA1000 chip to check size and integrity of components (Agilent). Double-stranded cDNA was labelled with Cy3 nonamers using 'Single Colour Labelling kit' (Nimblegen) and hybridised to custom-designed arrays (Nimblegen). As each slide contains three arrays the treatments were

distributed over the two slides (i.e. array 1 = two replicates of treatment 1 and one replicate of treatment 2) to minimise systematic variation. Amplified N-ChIP (H3K36me3 N-ChIP) and MeDIP samples were labelled with Cy5 or Cy3 nonamers using 'Two colour labelling kit' (Nimblegen). One of the three replicates included a dye swap to minimise variation introduced by dye biases. Labelled input and IP samples were consolidated and hybridised to custom-designed arrays (Nimblegen). For H3K27me3 and H3K4me3 N-ChIP DNA, samples were amplified using WGA2 (Sigma) and sent to B. Ylstra at the microarray facility of the VU University Medical Centre/ Cancer Centre Amsterdam for labelling, hybridisation and scanning of arrays. The microarrays used are detailed in table 2.2. Details of the tiled regions covered on the custom tiling array are provided in appendix 1. Microarrays were washed using 'Microarray Wash kit' (Nimblegen) and scanned using an MS200 scanner (Nimblegen) according to the manufacturer's instructions. Array images were processed and data was outputted using 'Nimblescan' software (Nimblegen).

Microarray	Format	Regions covered	Details
Mouse ChIP-chip 3x720K RefSeq Promoter Array (Nimblegen design)	HX3; 3 arrays on one slide; 720K probes each array	20,404 RefSeq promoters; -3200bp to +800bp relative to TSS on average	Median probe spacing = 100bp; Probe length = 50-75mer; Build = MM9
Custom tiling array ("Hox-plus")(Designed by Wendy Bickmore's lab)	HX3; 3 arrays on one slide; 720K probes each array	80 tiled regions; 556 gene promoters; includes all 4 extended *Hox* clusters; enriched for developmental genes and PcG targets (see appendix 1 for details)	Tiled (overlapping) probe layout for high resolution; 4 replicates of each probe on each array for robust measurements; Probe length = 50-75mer; Build = MM9

Table 2.2 - Nimblegen microarrays used for this project

2.8. Bioinformatic analyses and data processing

In general, the R language (http://www.r-project.org/)(version 2.12.1) and added packages were used for most methods in this section.

2.8.1. Data acquisition

In general data from published work was acquired through the Gene Expression Omnibus (GEO)(http://www.ncbi.nlm.nih.gov/geo/) or directly from supplementary data. *Dnmt1*$^{+/+}$/*Dnmt1*$^{-/-}$ MEF and *Dnmt1*$^{+/+}$ + 5-aza dC gene expression microarray data was inherited from a previous PhD student in the lab, Jamie Hackett. This data was generated using 'MouseWG-6 v2.0 BeadChip' array (Illumina) and had been processed and normalised by J. Hackett (Hackett *et al.*, in submission). Promoter DNA methylation data from MEFs was downloaded as a supplementary file with Meissner *et al.*, (2008) and was generated using reduced representation bisulfite sequencing. Promoter K4me3/K27me3 and CpG density was downloaded as a supplementary file with Mikkelsen *et al.*, (2007), histone marks were defined using ChIP-seq and promoter CpG density was defined according to Weber *et al.*, (2007). This gene level data was integrated with my data by merging on official gene symbol.

2.8.2. Statistical tests and plots

Fisher's exact and Chi-squared tests were used to assess distribution of samples among categories where there are two categories and >2 caetgores respectively, and were both performed using 'GraphPad Quick Calcs' (http://www.graphpad.com/quickcalcs/index.cfm). Mann-Whitney-U (MWU) test (also known as Wilcoxon-Rank-Sum test) is a non-parametric test that was used to test if two independent samples are equally large, and was performed using R. MWU test was used where sample size was sufficiently large and a normal distribution of samples could not be assumed. For small numbers of samples or where normality could be assumed a two-tailed Student's T-test was performed to test if independent samples are equally large. When assessing the correlation between two sets of paired samples a Pearson correlation test was performed using R or only the R-value (Pearson correlation co-efficient) or R^2-value was calculated. For calculating the probability

that a single value is part of a defined normal distribution the *pnorm* function in R was used. Gene ontology (GO) analysis was performed using 'Babelomics 4.2' (http://babelomics.bioinfo.cipf.es/), specifically the 'FATIGO' program. This program performs a two-tailed Fisher's exact test on the GO terms from two lists of genes that are provided by the user. As multiple hypotheses are questioned, the p-value is corrected by the false discovery rate (FDR) procedure of Benjamini and Hochberg. Bar and line charts were drawn in R and Microsoft Excel with error bars representing standard error of the mean. Bar plots showing microarray data at the probe level were generated in the UCSC browser (http://genome.ucsc.edu/). Scatter plots were drawn in R and Microsoft Excel. Boxplots were drawn in R using the default settings for whisker length and box height: Box vertical extremities represent upper and lower quartiles so the height of the box is the inter-quartile range (IQR). Black line within the box is the median. Upper and lower whiskers are defined as the upper quartile plus 1.5 times the IQR and lower quartile minus 1.5 times IQR respectively (default method). The whiskers were set as the sample maximum and minimum if these values are smaller or larger, respectively, than the values calculated using the default method above. Points that lie outside of the whiskers (outliers) were not plotted. Heatmaps were 'manually' drawn in R using the *image* function. For clustering of genes in R, the Euclidean distance between all genes was calculated and hierarchical clustering was performed using the 'Ward' method.

2.8.3. Genome annotations and annotation of microarray probes

Genome annotations were all derived from the *Mus musculus* (mm9) build (UCSC). The locations of genes, gene clusters, exons, introns and transcription start sites were all taken or derived from the 'RefGene_mm9' table downloaded from the UCSC browser. The locations of two types of CpG islands were annotated. One type is the CpG islands defined by the UCSC browser by sequence characteristics and these were defined based on a download from the UCSC browser. The other was experimentally determined in Illingworth *et al.*, (2008) and was derived from a supplementary file of the publication. Genes were deemed to be in 'tandem' if they appear next to each other in the 'RefGene_mm9' table and are on the same chromosome. Microarray probes were annotated based on their overlap with genomic features. A probe that shows any overlap with any part of a feature was annotated with that feature. For example, genic probes were defined as overlapping any transcript in the 'RefGene_mm9' table and the rest of the probes were annotated as 'intergenic'. The transcriptionally 'active' and 'inactive' portions of the *Hox* clusters were determined

arbitrarily based on cDNA array data. Promoter-associated probes on the Hox-plus custom tiled array were defined by any overlap with a tiled region covered by the promoter array and not overlapping a transcription start site or gene. This left probes that were up to 3.5kb upstream of a gene but were not overlapping the gene body.

2.8.4. cDNA tiled microarray data processing and analysis

Data from the six arrays for cDNA (three replicates of each treatment) were normalised using the Robust Multichip Average (RMA) method (Bolstad *et al.*, 2003) within 'Nimblescan' software (Nimblegen). RMA is commonly used for normalisation of expression arrays and works by subtracting background then adjusting the scale of data by quantile normalisation, manipulating the distribution of the data so that the quantiles match between samples (Bolstad *et al.*, 2003). This reduces systematic variation introduced by using multiple arrays. Subsequent data analysis was performed using R. As the Hox-plus custom arrays contain four copies of each probe, the mean of probe replicates was calculated and used in subsequent analyses. To define probes that have significantly higher or lower signals between treatments a two-tailed Student's T-test was performed for every probe between the treatments. Probes that have a $p<0.05$ and are in groups of three consecutive probes with $p<0.05$ were defined as different. This spatial method was preferred to a multiple testing approach for p-value correction as it will remove spurious signals produced by single probes, an important factor when considering data from tiled probes. Also, as a large number of hypotheses were tested, multiple testing correction may result in a lower number of probes making a certain cut-off. The false discovery rate (FDR) of the 3 consecutive probe method was estimated by permutation of sample replicates. The estimated FDR before the 3 consecutive probe method was very high at 46.5% but was reduced to 0.5% following application of the method, suggesting that it is reliable. For visual representation of cDNA array data the mean value for each probe was calculated for the three technical replicates. To reduce noise in the visual representations, the values were smoothed by taking the median of the signal from 5 probes and applying it to the central probe using a sliding window (*zoo* package in R). These probe values were visualised in the UCSC browser.

2.8.5. Two colour microarray data processing and analysis

Processing of two-colour microarray data was performed in R using modified scripts that were originally written by Tobias Straub (LMU, Munich) for analysis of protein-DNA interactions by ChIP-chip (downloaded from http://genome1.bio.med.uni-muenchen.de/downloads.htm). Data processing included quality control, intra-array normalisation and inter-array normalisation. The first step was quality control. For arrays

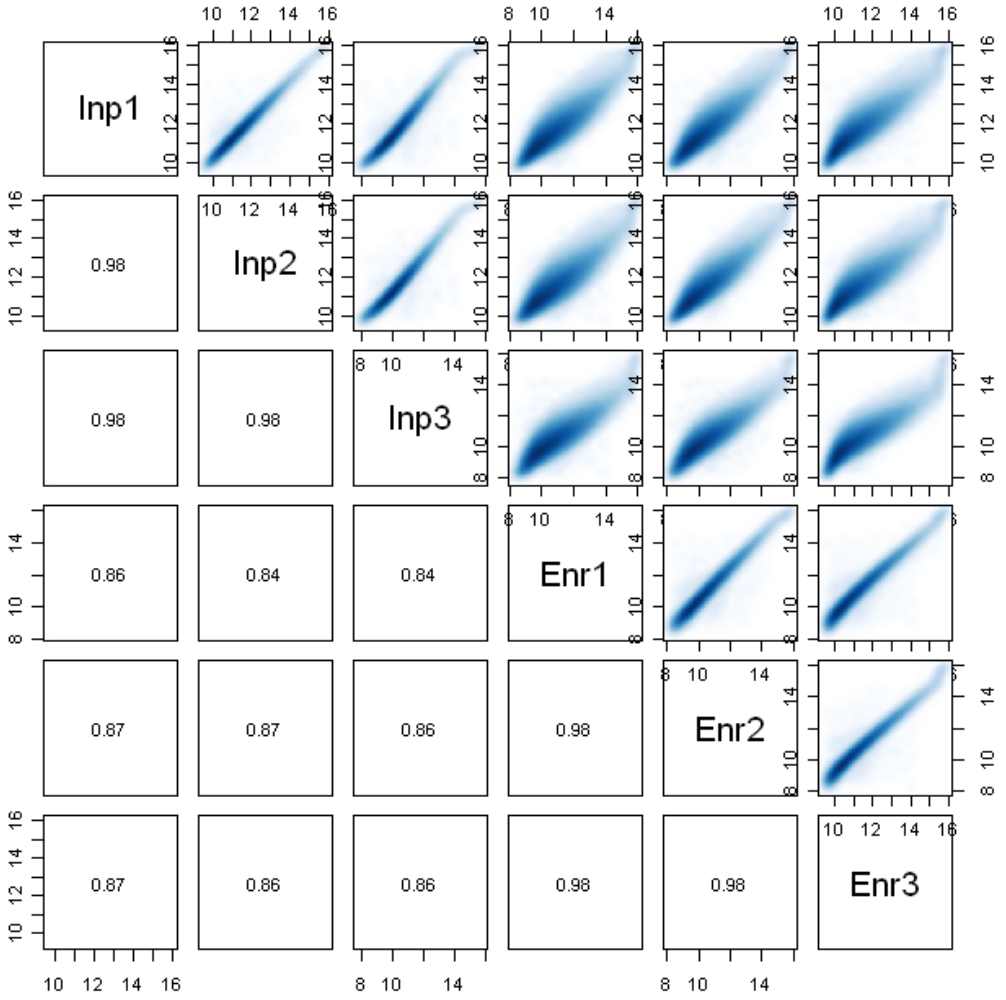


Figure 2.1 - Correlations between array signals. Scatter plots (upper right) and Pearson correlation co-efficients (lower left) were produced for all array signals. Note good correlations within input samples and enrichment samples but not between them. Data here is from MeDIP experiment and is used as an example although all two-colour arrays performed produced similar relationships.

scanned in-house, the array images were checked for artefacts such as scratches and hybridisation artefacts. For arrays performed by the service centre, array images were recreated from the data files using R to check for artefacts. Correlations between array signals between replicates were measured by scatter plot (Fig. 2.1). This is an important step for two reasons. Firstly, it is important to establish that the raw data is of good quality as no normalisation can account for unsatisfactory raw data. 'Good quality' data entails a good linear correlation between replicate input signals and between replicate enrichment signals but less dependence between input and enrichment values. Secondly, the relationship between replicates is important for deciding which type of normalisation to apply (see 'inter-array normalisation' below).

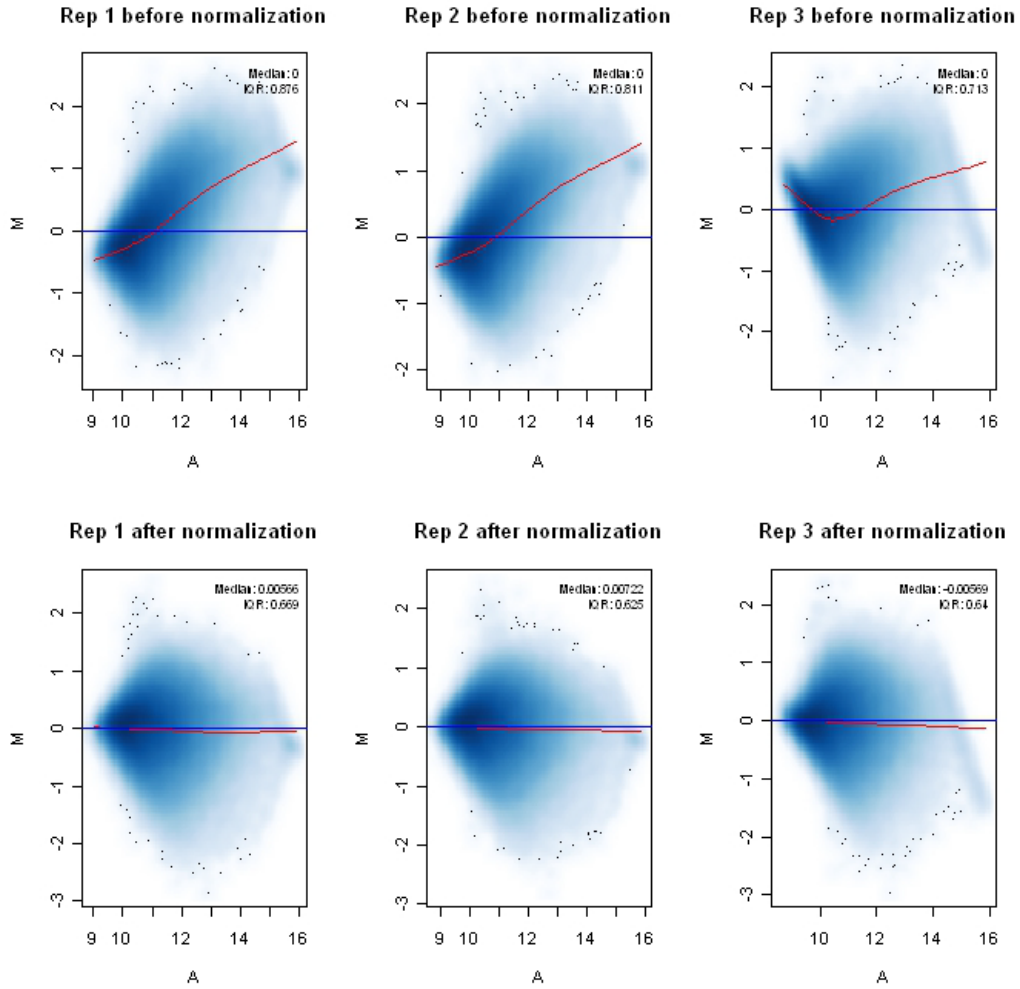


Figure 2.2 - Data before and after intra-array normalisation by lowess. MA-plots show the dependence of log2(IP/INP) signal on overall signal strength. M = log2(IP/INP), A = mean of IP and INP. Red line is fit by loess. Data here is from MeDIP experiment and is used as an example although all two-colour arrays performed produced similar relationships.

The first normalisation that was applied is to calculate the log2 ratio of enrichment/immunoprecipitated signal (IP) and input signal (INP)(log2(IP/INP)) and to scale these values around 0. This controls for many variables such as copy number variation. Next, intra-array normalisation was performed to correct for differences in kinetics of the two dyes used. MA-plots can be used to check for dependence of the log2(IP/INP) signal on overall signal strength, a signature of dye bias (Fig. 2.2). Dye bias is often observed with two-colour arrays (Choi and Do, 2006) and this was the case for my data, as indicated by a positive relationship on the MA-plots (Fig. 2.2 upper panels). The relationship displayed on the plot for replicate 3, which was a dye swap, is different from the others further supporting the notion that dye bias causes this effect. Locally weighted scatter plot smoothing (lowess) is a standard method used to correct for dependence of M on A (Cleveland, 1979; Choi and Do, 2006). This method was applied to the data using the *limma* package in R, producing a much reduced dependence (Fig. 2.2 lower panels).

The final normalisation step was inter-array normalisation which reduces systematic variation caused by, for example, the use of multiple microarray slides or the effect of performing experiments at different times. As the replicate signals are well correlated throughout their distribution (assessed by the straight lines produced in Fig. 2.1 across the entire range) scale normalisation can be used for this step (Choi and Do, 2006). Scale normalisation acts to adjust the distribution of samples so that they align with each other (Fig. 2.3). Scale normalisation was performed using the *limma* package in R. Figure 2.3 shows that the overall signal strengths can differ between replicates and that this can be corrected using scale normalisation.

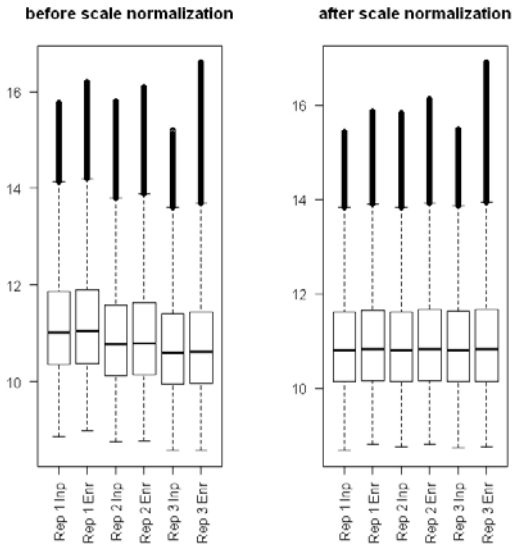


Figure 2.3 - Data before and after inter-array normalisation. Boxplots of the log2 transformed signal intensities are shown for each replicate. Note that after normalisation the medians of each sample are aligned. Data here is from MeDIP experiment and is used as an example although all two-colour arrays performed produced similar relationships.

For visualisation of data the mean of the three replicates was calculated from the normalised data and visualised in the UCSC browser. For visualisation of MeDIP data, a smoothing sliding window approach was used to find the median of 9 consecutive probes and apply it to the central probe using the *zoo* function in R. For H3K27me3 and H3K4me3 promoter N-ChIP, data was analysed at the promoter level. For this, the normalised mean probe signal (log2(IP/INP)) within the entire promoter tiled region for each gene was calculated and the difference between treatments was computed. The distribution of these difference values for every promoter region was plotted and, as the number of promoters that were different between treatments was large, arbitrary cut-offs were used. The upper and lower cut-offs were defined as the median plus or minus 4 times the median absolute deviation (MAD) respectively.

For H3K36me3 data, as the Hox-plus tiling arrays contain 4 replicates for each probe, the mean was calculated. For defining probes that show significantly different H3K36me3 signals between treatments a two-tailed Student's T-test was applied for each probe. Significantly different probes were defined as having $p<0.05$ and being in a group of 3 consecutive probes each with $p<0.05$. This was performed to reduce noise introduced by spurious signals of single probes. An estimation of the FDR for this method by permutation was 0.4%.

2.8.6. HELP-tag-seq data processing and analysis

HpaII tiny fragment Enrichment by Ligation-mediated PCR assay coupled with tag sequencing (HELP-tag-seq) was performed by the lab of J.M.Greally as part of a collaboration (Khulan *et al.,* 2006; Oda *et al.,* 2009; Suzuki *et al.,* 2010). This form of the assay is still in development so the data processing steps that I performed are detailed here.

During the previous generations of the HELP protocol genomic DNA is subjected to parallel digests with the isoschizomers MspI (non methyl-sensitive) and HpaII (methyl-sensitive), the fragments are size fractionated, and the products are analysed by microarray or sequencing (Khulan *et al.,* 2006; Oda *et al.,* 2009). This lead to the requirement for two HpaII cuts within a certain distance from each other for a site to be represented in the HpaII library. The MspI digest controls for biases that can arise in PCR amplification or sequencing (Oda *et al.,* 2009; Suzuki *et al.,* 2010). In this latest generation of the assay, HELP-tag-seq, a first adaptor is ligated to both ends of each fragment which contains an EcoP15l restriction site. This enzyme cuts in the flanking sequence allowing a defined sequence 'tag' to be generated for each MspI/HpaII cut, meaning that no size fractionation is

performed (Suzuki *et al.*, 2010). This method was performed on *Dnmt1*$^{+/+}$ MEF genomic DNA and the 'tags' were Illumina sequenced and mapped back to the genome by the lab of J.Greally. The number of HpaII reads relative to the number of MspI reads provides a measure of 5mC (and technically 5hmC) against unmodified cytosine at each CCGG site analysed. The relationship between HpaII tags and MspI tags was measured by the angle produced between the x-axis and a straight line formed by connecting the graph origin to the point produced by plotting the number of MspI reads against the number of HpaII reads for a given CCGG site (Fig. 2.4-left)(as suggested by Suzuki *et al.*, 2010). The angle gives an inverse measure of methylation as methylated sites would have less HpaII reads relative to MspI reads and hence a low angle. The length of the line, 'C', formed between the point and the origin gives an indication of the overall number of reads for that CCGG site and hence can be used to score confidence of a particular angle (Fig. 2.4-left). I examined the effect of increasing the cut-off for 'C' (confidence threshold) on the number of CCGG sites that can be analysed (Fig. 2.4-right). At this sequencing depth, using any cut-off showed a strong effect on the number of CCGG that can be analysed, with ~200,000 sites remaining at a cut-off of C=10.

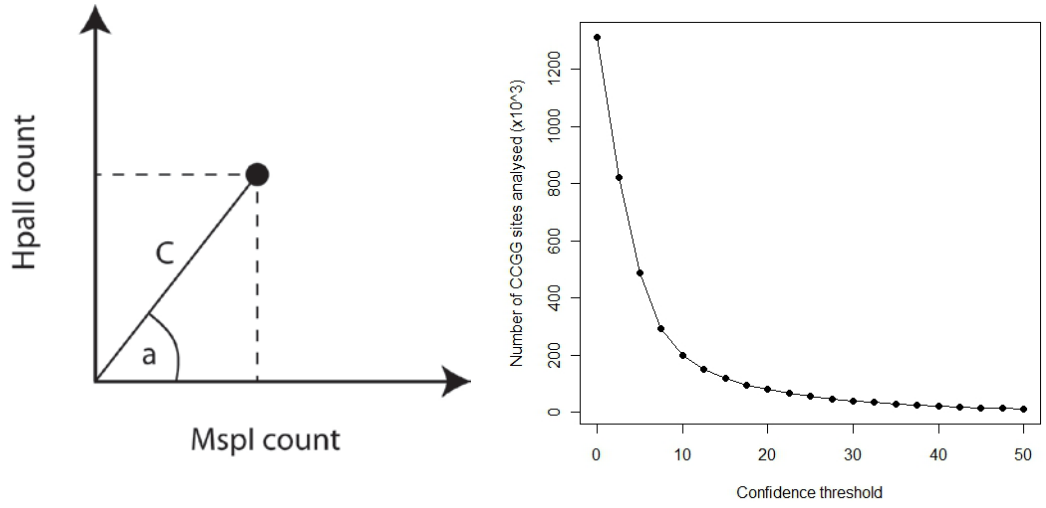


Figure 2.4 - Data processing for HELP-seq. (Left) The method used to measure relationship between MspI reads and HpaII reads. For each CCGG site the number of HpaII reads is plotted against the number of MspI reads and the angle a is calculated. The length of line C is calculated and represents a measure of sequence depth and hence confidence. (Right) The number of CCGG sites remaining at each cut-off of confidence level (C).

One might expect that at some sites where only a small number of reads were mapped that an artificially low angle would be observed due to the potential of HpaII reads representing a large proportion of total reads. For example, a site with only 1 sequence read

in the HpaII library and none in the MspI library will give a angle of 0 (100% methylation). This possibility was examined by plotting the number of high-angle (>70°, low methylation) and low-angle (<20°, high methylation) CCGG sites with increasing cut-offs for 'C' (Fig. 2.5). While the high-angle (low methylation) sites showed a steady decrease in number as 'C' was increased, the number of low-angle (high methylation) sites decreased rapidly at low cut-offs for 'C' suggesting that the prediction is true (Fig. 2.5). For this reason, the cut-off for 'C' was set at 10 to remove the large number of 'false' methylated sites while maintaining a maximum possible number of sites. As CCGG sites are enriched in CpG islands and CpG-rich regions of the genome, we would expect the majority of sites to be hypomethylated, and indeed this has been observed in previous studies (Brunner *et al.*, 2009). Using a cut-off of C=10 produces the expected bias towards hypomethylated sequences (Fig. 2.5).

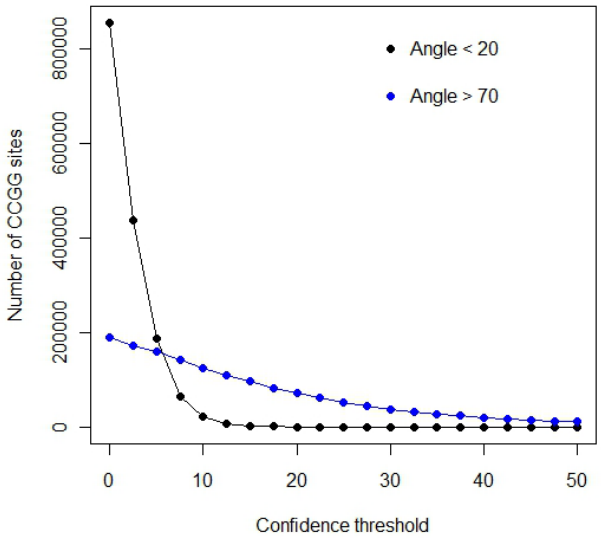


Figure 2.5 - The effect of increasing confidence thresholds on the number of high-methylated and low-methylated sites. The number of CCGG sites remaining at increasing cut-offs of confidence (C) for high-methylated (angle<20) and low-methylated (angle>70) CCGG sites.

The range of the angle values for all CCGG sites was stretched from 90 (0 - 90) to be a more intuitive 100 (0 - 100) giving an 'adjusted angle' (performed by J.M.Greally lab). I assigned a 'methylation score' (from 0-100) to each CCGG site by subtracting the 'adjusted angle' from 100 to make interpretation more intuitive. CCGG sites were annotated in the same way as microarray probes (described in section 2.8.3). CCGG sites were annotated as within a promoter if they were within -1000 and +500bp of a transcription start site in the 'RefGene_mm9' table (UCSC). Promoter level methylation was calculated by taking the mean and standard deviation of the methylation score of CCGG sites within promoter regions. Standard deviation between sites within a promoter region was generally low (data not shown) so promoters that contain only one CCGG site measured were included in promoter analyses.

As this assay is in development I felt it important to validate the results. I performed bisulfite sequencing on 16 CCGG sites in *Dnmt1*$^{+/+}$ MEF DNA from a range of methylation levels and compared the results to those achieved with HELP-tag-seq (Fig. 2.6). Although there was a clear correlation between the two methods they showed considerable variation, as indicated by an R^2 value of 0.24 (Fig. 2.6). The source of this variation is unclear. Approximately 18 sequences were analysed for each CCGG for bisulfite sequencing, meaning that some variation may arise here. Similar correlations were observed in published validations of HELP based methods with bisulfite sequencing (Thompson *et al.*, 2008; Brunner *et al.*, 2009; Suzuki *et al.*, 2010). A particularly poor correlation was observed between HELP-tag-seq data and both bisulfite conversion-based Illumina Infinium array and methyl-DNA immunoprecipitation data (Brunner *et al.*, 2009). It is currently unclear from where this variation arises. Considerable variation has been observed between technical replicates indicating that experimental noise may contribute to variability for this assay (Ball *et al.*, 2009). It is suggested that CCGG sites with low DNA methylation levels show greater noise using this assay (Ball *et al.*, 2009). One possibility is that decreased noise could be achieved with increased sequence depth, although this remains to be tested (Ball *et al.*, 2009). In conclusion, HELP-tag-seq data, at the sequence depth achieved here, should be interpreted with caution and, where possible, important data should be validated using an independent method.

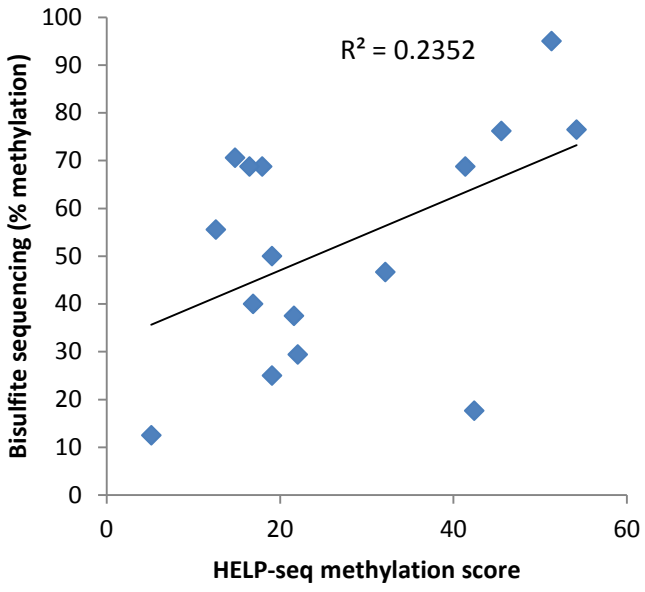


Figure 2.6 - Comparison of HELP-tag-seq and bisulfite sequencing at individual CCGG sites. Correlation is assessed by Pearson's R^2 value.

Chapter 3 - Derepression and loss of H3K27me3 at PcG target genes in DNA hypomethylated somatic cells

3.1. Introduction

It is widely accepted that DNA methylation affects transcription, but, how it does so is less clear. I first set out to investigate the pattern of gene mis-expression in experimentally hypomethylated cells, with the long-term aim of finding evidence for non-canonical functions of DNA methylation in gene regulation. A large number of genes are reportedly mis-expressed in mouse tissues and cell lines with mutations in DNA methyltransferases, suggesting that DNA methylation is an integral component of gene regulation in mammals (Jackson-Grusby *et al.*, 2001; Lande-Diner *et al.*, 2007; Fouse *et al.*, 2008; Sen *et al.*, 2010). An extensive cross-talk between DNA methylation and other epigenetic systems is emerging implicating DNA methylation in additional levels of chromatin organisation (Zemach *et al.*, 2010; Bartke *et al.*, 2010). Also, non-promoter functions of DNA methylation in gene regulation have been postulated suggesting that this epigenetic mark is more than the occasional repressor of gene promoters. These include a conserved but unknown function of DNA methylation at gene bodies and a putative interaction with the polycomb system of gene repression (Lister *et al.*, 2009; Wu *et al.*, 2010). The links with the PcG proteins are particularly interesting from a gene regulation standpoint as these proteins are thought to have a major role in shaping gene expression patterns. Therefore, in this chapter I focussed initially on the emerging link between DNA methylation and PcG repression. During the course of this chapter I uncovered an unexpected relationship between DNA methylation and PcG repression at classic polycomb targets the *Hox* genes.

3.2. Results

3.2.1. Dnmt1$^{-/-}$ MEFs are severely DNA hypomethylated

In order to study the role of DNA methylation on gene regulation and chromatin structure I required an appropriate experimental system. I chose to study mouse embryonic fibroblasts (MEFs) that are homozygous for a mutation in *Dnmt1*, the gene encoding the major maintenance DNA methyltransferase (Lande-Diner *et al.,* 2007). This mutation (*Dnmt1*$^{n/n}$) was introduced by targeted recombination by Li *et al.* (1992) and was designed to ablate *Dnmt1* coding potential. In fact this mutation appears to allow expression of a truncated form of Dnmt1 protein, albeit at greatly reduced levels, so is likely a severe hypomorphic allele (Li *et al.,* 1992). This mutation is embryonic lethal when homozygous at around 11dpc and viable MEFs cannot be generated (Li *et al.,* 1992; Lande-Diner *et al.,* 2007). This allele has been previously combined with homozygosity for a null allele of *Trp53* (*Trp53*$^{-/-}$), encoding the p53 protein, which allowed for the derivation of viable MEFs, probably due to suppression of apoptosis (Jackson-Grusby *et al.,* 2001; Lande-Diner *et al.,* 2007). *Trp53*$^{-/-}$, *Dnmt1*$^{n/n}$ MEFs are therefore compared to *Trp53*$^{-/-}$, *Dnmt1*$^{+/+}$ sibling matched mouse embryonic fibroblast cells as control (Lande-Diner *et al.,* 2007). For brevity, the genotype of these cell lines will be referred to as *Dnmt1*$^{+/+}$ (*Trp53*$^{-/-}$, *Dnmt1*$^{+/+}$) and *Dnmt1*$^{-/-}$ (*Trp53*$^{-/-}$, *Dnmt1*$^{n/n}$) herein. There are a number of advantages to studying these cells. They have been suggested to exhibit severe DNA hypomethylation yet are viable in culture (Lande-Diner et al., 2007; Hackett, 2010, PhD thesis). To my knowledge they represent the only long term viable somatic cells with a severe mutation in *Dnmt1*. Mouse ES cells are of course viable in the undifferentiated state lacking detectable DNA methylation, but initiate apoptosis upon differentiation (Jackson *et al.,* 2004). Importantly, despite *Trp53* mutation these cells are thought to be non-transformed as they do not grow in three dimensional culture in matrigel (Hackett, 2010, PhD thesis). These cells therefore represent a valuable model to study the role of DNA methylation in somatic cells as they are severely hypomethylated, non-transformed and viable in culture (Lande-Diner et al., 2007; Hackett, 2010, PhD thesis). Despite the advantages, I was also aware of the disadvantages of this cell line system, including the potential variation introduced by the heterogeneity of MEFs. Therefore, I decided to supplement studies using these cells with use of a small molecule inhibitor of

DNMTs on cultured cells, and also by studying mouse embryos with *Dnmt1* mutation (Chong *et al.*, 2007). These experiments will be described later in the chapter.

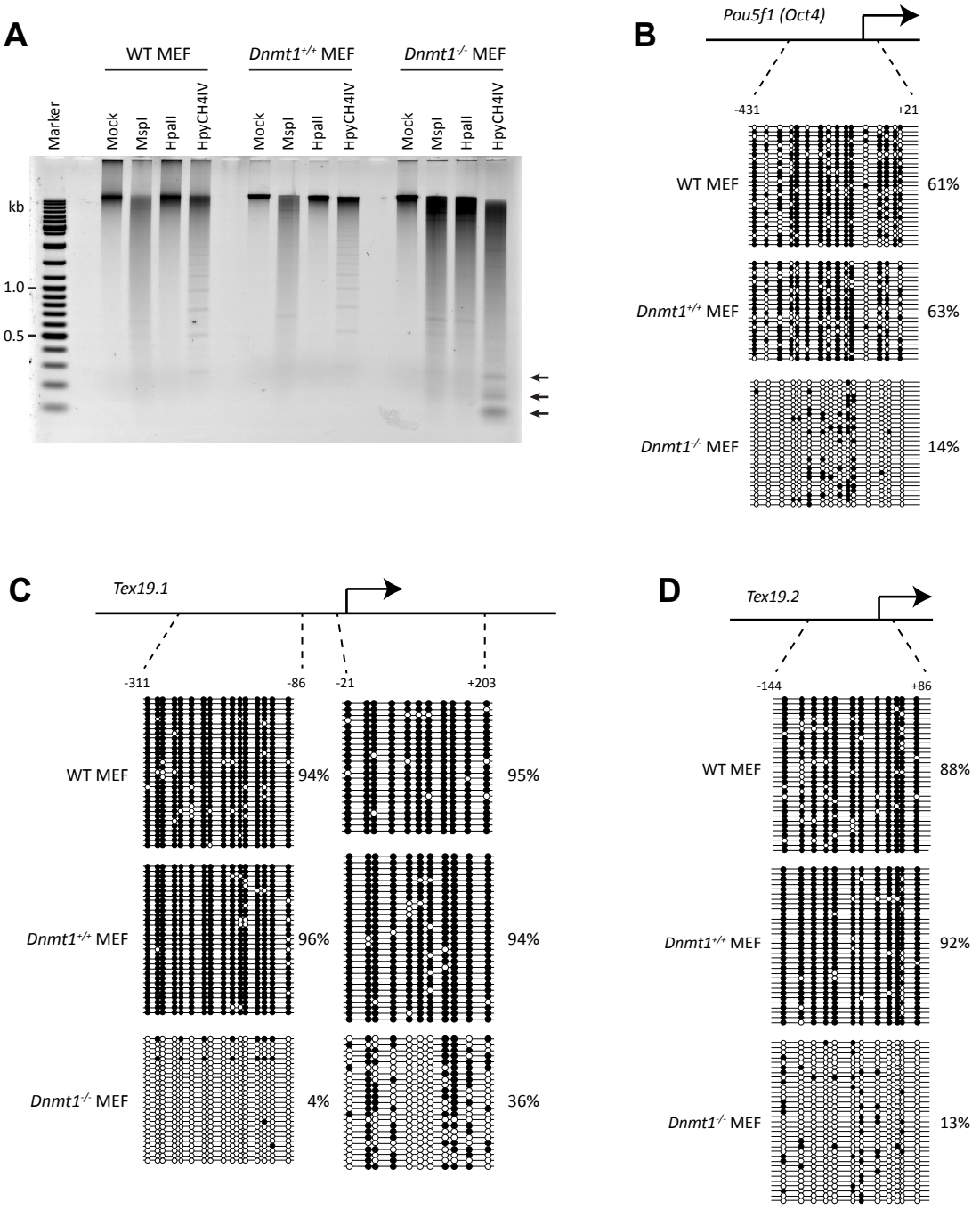
I began by confirming that *Dnmt1*$^{-/-}$ MEFs are indeed DNA hypomethylated compared to controls. Firstly I performed methylation-sensitive restriction digest on genomic DNA isolated from these cells along with DNA from *Dnmt1*$^{+/+}$ MEFs and primary wildtype MEFs (Fig. 3.1A). Increased smearing and the appearance of smaller molecular weight fragments was observed for *Dnmt1*$^{-/-}$ MEF DNA using methylation sensitive endonucleases HpaII and HpyCH4IV, indicating hypomethylation at CCGG and ACGT sites (Fig. 3.1A). The appearance of low molecular weight fragments (marked by arrows) in HpyCH4IV digest of *Dnmt1*$^{-/-}$ MEF DNA is consistent with hypomethylation of gamma satellite DNA (Fig. 3.1A; Abdurashitov *et al.*, 2009). HpaII digested DNA from *Dnmt1*$^{-/-}$ MEFs produced a similar pattern to that digested with its methylation insensitive isoschizomer, MspI, suggesting severe hypomethylation at these sites. To quantify hypomethylation at certain loci I performed bisulfite sequencing for the promoter regions of *Pou5f1*, *Tex19.1* and *Tex19.2* which have been shown to be highly methylated in somatic tissues (Hackett, 2010, PhD thesis). All three of these loci showed greater than 60% less CpG methylation in *Dnmt1*$^{-/-}$ MEFs relative to controls (Fig. 3.1B to D). By restriction digest and bisulfite sequencing of selected loci, *Dnmt1*$^{+/+}$ MEFs and primary WT MEFs showed no gross difference in DNA methylation status indicating that *Trp53* mutation does not cause widespread changes in DNA methylation (Fig. 3.1A to D). Together, these data suggest that *Dnmt1*$^{-/-}$ MEFs are severely DNA hypomethylated, at least at the loci and sites assessed here, and therefore provide an adequate resource for studies of the roles of DNA methylation.

3.2.2. Many genes upregulated in experimentally hypomethylated cells are PcG targets

I reasoned that studying the characteristics of genes that are upregulated in DNA hypomethylated cells may help to elucidate the mechanisms by which DNA methylation facilitates transcriptional repression. I was particularly interested in the histone modifications that these genes are associated with as I postulated that it may be possible to uncover novel links between DNA methylation and chromatin structure using this approach. My starting point was expression microarray data for *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs that I inherited from J.Hackett, a previous PhD student in the lab. He used these cells, among

Figure 3.1 - *Dnmt1*$^{-/-}$ mouse embryonic fibroblasts are severely DNA hypomethylated

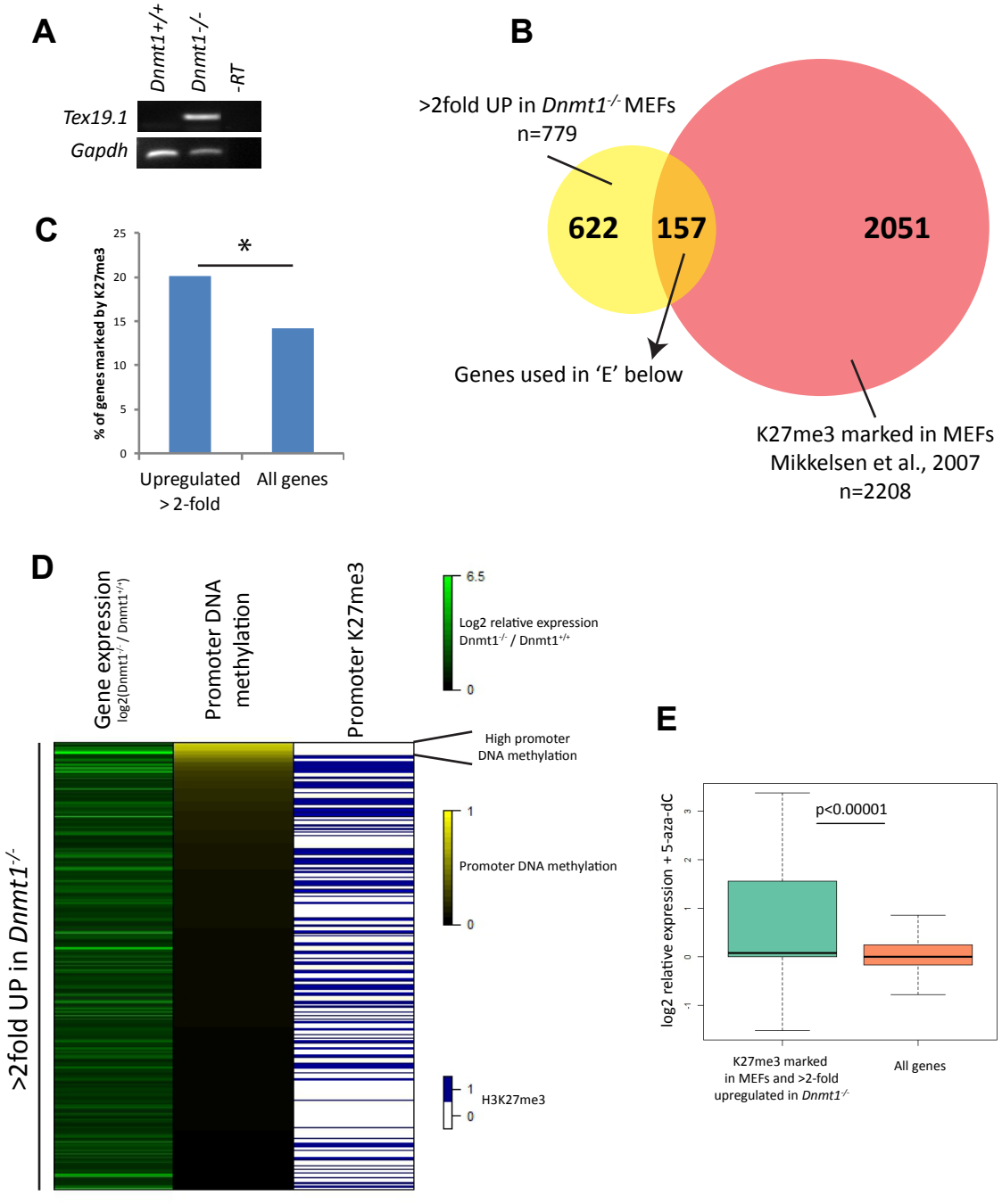
A. Digest of genomic DNA from wildtype primary mouse embryonic fibroblasts (WT MEF), *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs with restriction endonucleases. MspI and HpaII are isoschizomers that cut at CCGG with HpaII being inhibited by CpG methylation and MspI not. HpyCH4IV cuts at ACGT and is blocked by CpG methylation. The presence of low molecular weight DNA for *Dnmt1*$^{-/-}$ MEFs is indicative of hypomethylation at these sites. Arrows mark the position of low molecular weight bands visible in HpyCH4IV digest of *Dnmt1*$^{-/-}$ MEF DNA indicative of hypomethylation at gamma-satellite DNA (Abdurashitov *et al.,* 2009). B to D. Bisulfite sequencing for gene promoter regions that are thought to be highly DNA methylated in somatic tissues. Filled circles represent methylated CpGs while unfilled circles represent unmethylated CpGs. The location of each region in relation to the transcription start site is shown. The percentage of CpGs within the region that are methylated is shown.



other experimentally hypomethylated cells, to identify genes that are expressed in the germline that utilise promoter DNA methylation for repression in somatic lineages (Hackett *et al.,* in submission). I confirmed that one such gene, *Tex19.1,* is de-repressed in *Dnmt1*$^{-/-}$ MEFs by RT-PCR, serving as a positive control (Fig. 3.2A). In total, 779 genes were identified as being >2-fold upregulated in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs (Fig. 3.2B). This list contained many germline expressed genes, genes on the X chromosome and imprinted genes consistent with current knowledge (Li *et al.,* 1993; Jackson-Grusby *et al.,* 2001; Lande-Diner et al., 2007; Borgel *et al.,* 2010; Velasco *et al.,* 2010). I noticed that among the upregulated genes were multiple genes from the *Hox* clusters and other classic targets of the polycomb repression system that are normally marked by the PRC2-associated K27me3 histone mark in a variety of tissues, including MEFs (Mikkelsen *et al.,* 2007). In order to characterise this observation in more detail, I used published K27me3 ChIP-seq mapping data from Mikkelsen *et al.* (2007) to ask how many of the upregulated genes are normally K27me3 marked in MEFs. I found that 157 of the 779 (about 20%) upregulated genes are K27me3 marked at their promoter regions in MEFs according to Mikkelsen *et al.* (2007)(Fig. 3.2B). This represents a modest but highly statistically significant enrichment over what would be expected by chance as <15% of all genes are marked by K27me3 in MEFs according to this study ($p<0.0001$, Fisher's exact, Fig. 3.2C). This observation was initially puzzling as dense DNA methylation and the PcG-mediated K27me3 mark have been suggested to be exclusive modifications (Lister *et al.,* 2009; Hawkins *et al.,* 2010; Lindroth *et al.,* 2008). In order to assess the promoter DNA methylation level of these genes I obtained published promoter DNA methylation data for MEFs that was generated by reduced representation bisulfite sequencing from Meissner *et al.* (2008). Taking the genes that overlap between the three datasets I plotted a heatmap of relative expression in *Dnmt1*$^{-/-}$ MEFs, promoter DNA methylation and K27me3 status of genes that are >2-fold upregulated in *Dnmt1*$^{-/-}$ MEFs (Fig. 3.2D). This plot showed that the following is true for these overlapped genes: 1) Only a small proportion have high levels of promoter DNA methylation in MEFs according to Meissner *et al.* (2008). Consistent with previous observations regarding dense CpG methylation and the K27me3 mark, these genes are not marked by K27me3 (indicated on Fig. 3.2D). 2) There is no obvious correlation between promoter DNA methylation level and relative expression in *Dnmt1*$^{-/-}$ MEFs. 3) Many of these genes do not have high promoter DNA methylation and instead are marked by the PRC2-associated K27me3 mark in MEFs (Fig. 3.2D). This indicates that many genes upregulated in *Dnmt1*$^{-/-}$ MEFs are not 'canonical' DNA methylation regulated genes but may be targets of the PcG proteins. As one would expect many secondary gene expression

Figure 3.2 - Many genes upregulated in hypomethylated MEFs are marked by K27me3

A. RT-PCR showing de-repression of the *Tex19.1* gene in *Dnmt1*$^{-/-}$ MEFs. This gene was used as a positive control as it is suggested to require promoter DNA methylation for repression in somatic cells (Hackett et al., in submission). B. Many genes that are >2-fold upregulated in *Dnmt1*$^{-/-}$ MEFs are marked by promoter K27me3 in MEFs. C. Proportion of upregulated genes, and all genes, that are marked by K27me3 in MEFs. * p<0.0001, Fisher's exact. D. Heatmap showing relative expression as determined by microarray (left panel), promoter DNA methylation in MEFs (middle panel) and promoter K27me3 status in MEFs (right panel) of genes upregulated > 2-fold in *Dnmt1*$^{-/-}$ MEFs. Genes (rows) are ordered by promoter DNA methylation. E. Boxplot showing the relative expression of genes that are both >2-fold upregulated in *Dnmt1*$^{-/-}$ MEFs and marked by H3K27me3 in MEFs, in *Dnmt1*$^{+/+}$ MEFs treated with demethylating agent 5-aza-dC (AZA). P-value shown was generated from a Mann-Whitney U test. Gene promoter K27me3 and DNA methylation data from MEFs used in this figure was taken from Mikkelsen et al., (2007) and Meissner et al., (2008).



changes in *Dnmt1*$^{-/-}$ MEFs, this observation by itself is of course not conclusive. However, given the links between DNA methylation and PcG proteins, described in chapter 1, I decided to investigate this further.

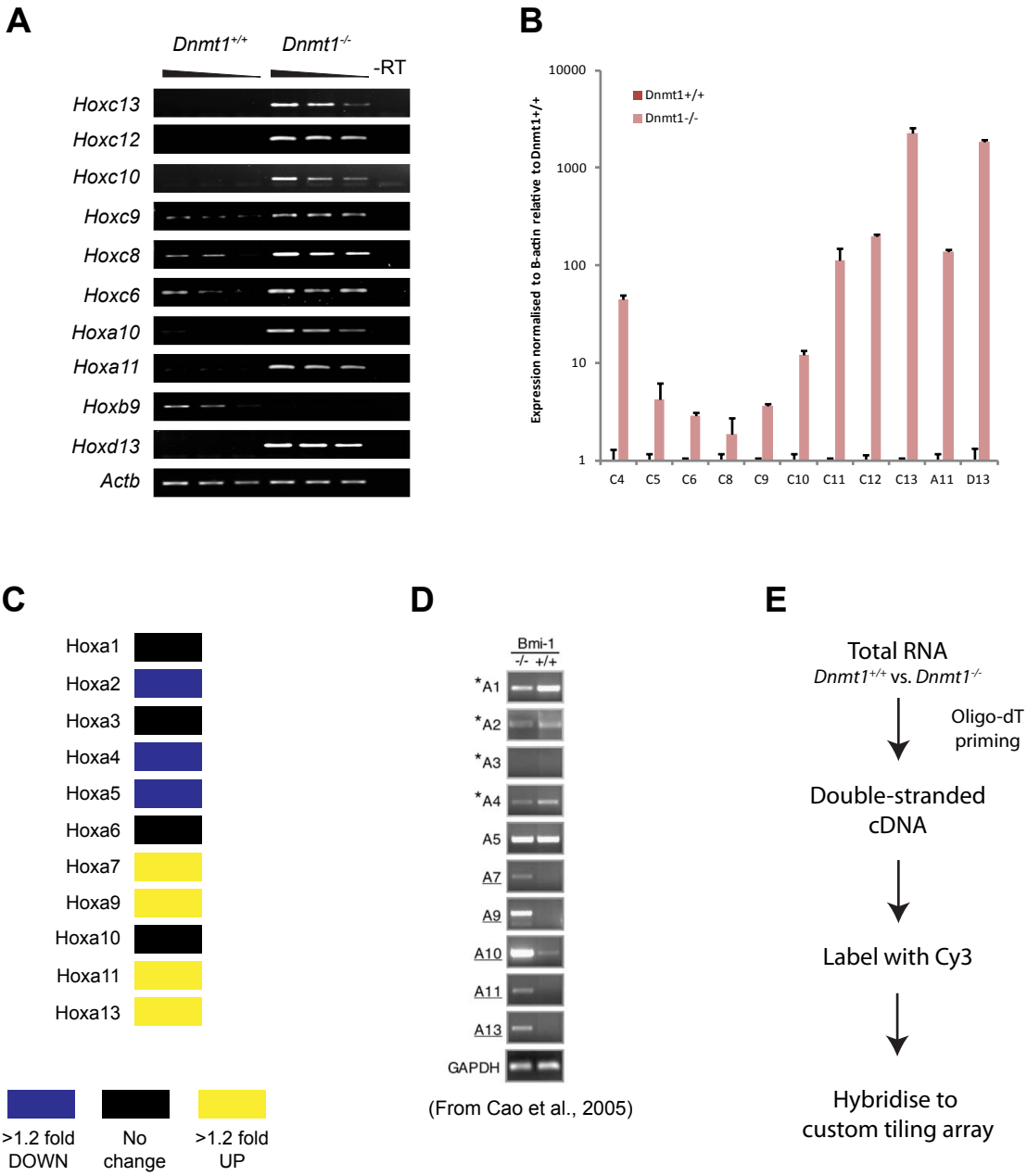
When using cell lines to study gene expression, particularly MEFs which are originally a heterogeneous population, it is often essential to use another system in parallel. This was particularly necessary in my case when studying *Hox* genes and other PcG target genes that show highly tissue-specific expression patterns. For this reason, expression microarray data from *Dnmt1*$^{+/+}$ MEFs treated with demethylating agent 5-aza-2'-deoxycytidine (5-aza-dC) was also studied (data inherited from J. Hackett, Hackett *et al.*, in submission). 5-aza-dC is an analogue of cytosine that is incorporated into DNA of dividing cells. DNMTs are irreversibly bound by 5-aza-dC leading to the degradation of bound DNMTs, but also of free DNMTs, causing passive loss of DNA methylation over cell divisions (Patel *et al.*, 2010). I found that, as a group, genes that are K27me3 marked at their promoter in MEFs and >2-fold upregulated in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs are upregulated in 5-aza-dC treated *Dnmt1*$^{+/+}$ MEFs by expression microarray (Fig. 3.2E, see Fig. 3.2B for gene set used). Importantly, although all of these genes were not observed to be upregulated upon 5-aza-dC treatment, upregulation was observed for certain tissue-specific genes (confirmed in a later figure). Together, these data suggest that many tissue-specific PcG target genes are de-repressed in DNA hypomethylated MEFs.

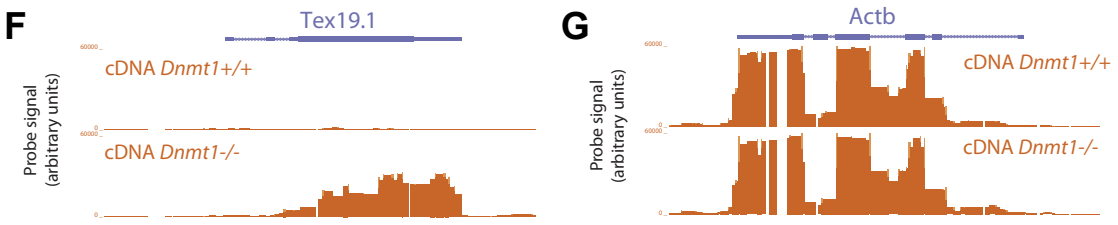
3.2.3. The 'Hox code' is perturbed in Dnmt1$^{-/-}$ *MEFs*

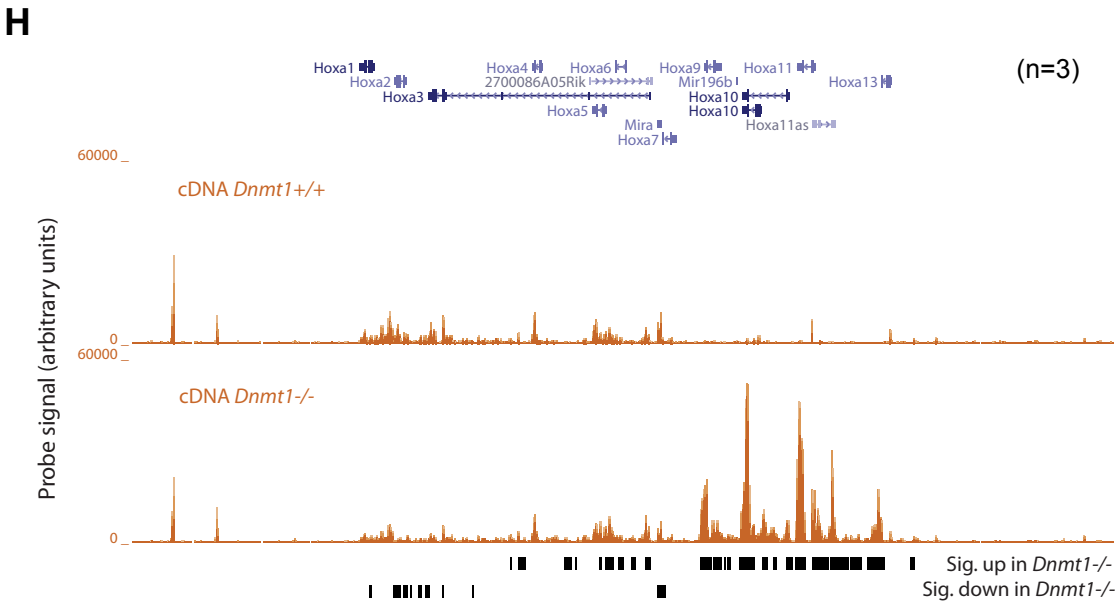
In order to investigate the putative link between DNA methylation and PcG repression in more detail I concentrated on the *Hox* genes, as many of these genes showed upregulation in *Dnmt1*$^{-/-}$ MEFs by microarray and they represent well characterised PcG target genes (Soshnikova and Duboule, 2009a). I performed RT-PCR and qRT-PCR on selected *Hox* genes from all four genomic clusters in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs (Fig. 3.3A and B). These results confirmed findings from the expression microarray but also showed that the number of *Hox* genes that are mis-regulated in *Dnmt1*$^{-/-}$ MEFs is larger than that suggested by microarray. The position of *Hox* genes within their cluster is thought to be important for their regulation as tissues and cells often show sub-domains of *Hox* clusters that share a transcriptional state (Soshnikova and Duboule, 2009b; Rinn *et al.*, 2007). For this reason I examined the expression of *Hox* genes in relation to their position within their clusters. The *HoxA*, *HoxD* and *HoxC* clusters showed evidence of a cluster location effect as *Hox9-13* genes were strongly de-repressed (Fig. 3.3A and B). In fact, the *HoxA* cluster showed

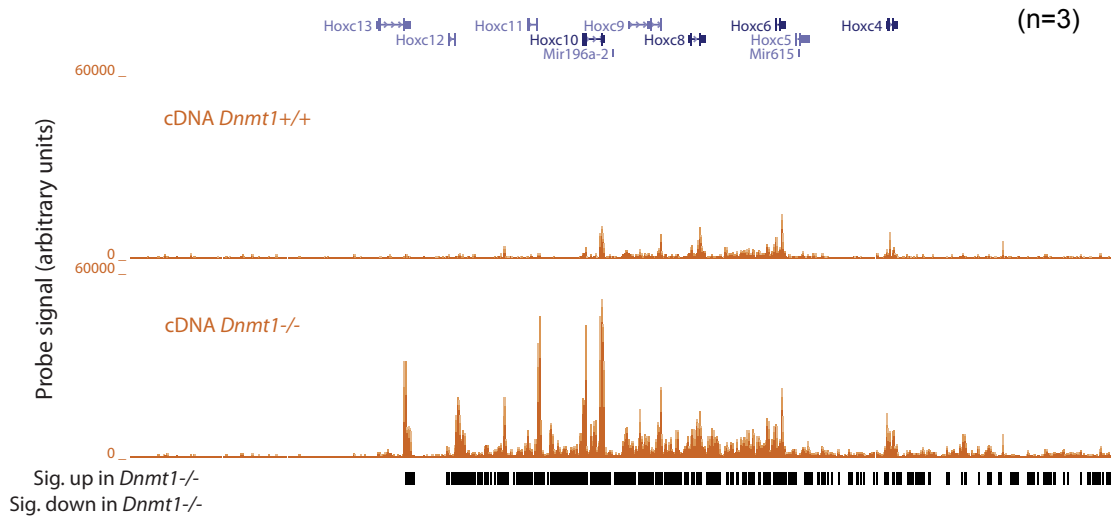
Figure 3.3 - Altered *Hox* gene expression in *Dnmt1*$^{-/-}$ MEFs

A. RT-PCR for *Hox* genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. A dilution series of cDNA was used. -RT is a no reverse transcriptase control. B. qRT-PCR for *Hox* genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars represent standard error of the mean of technical replicates. C. Summary of gene expression from expression microarray data (as described in previous figure) for the *HoxA* cluster in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs. D. Figure taken from Cao et al. (2005) showing RT-PCR for *HoxA* genes in *Bmi1*$^{+/+}$ and *Bmi*$^{-/-}$ MEFs. * and underlined indicates down-regulation and up-regulation respectively in *Bmi*$^{-/-}$ relative to *Bmi1*$^{+/+}$ as interpreted by Cao et al. E. Schematic of cDNA-custom tiling array experiment. F and G. cDNA signal from custom tiling microarray at control genes for *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Arbitrary units on the y-axis for probe signal. Position of RefSeq genes is shown above in blue. H to K. As 'F' and 'G' but for the four Hox gene clusters. Probes that are significantly higher or lower in Dnmt1-/- relative to Dnmt1+/+ MEFs are indicated by black bars below (see methods for details).

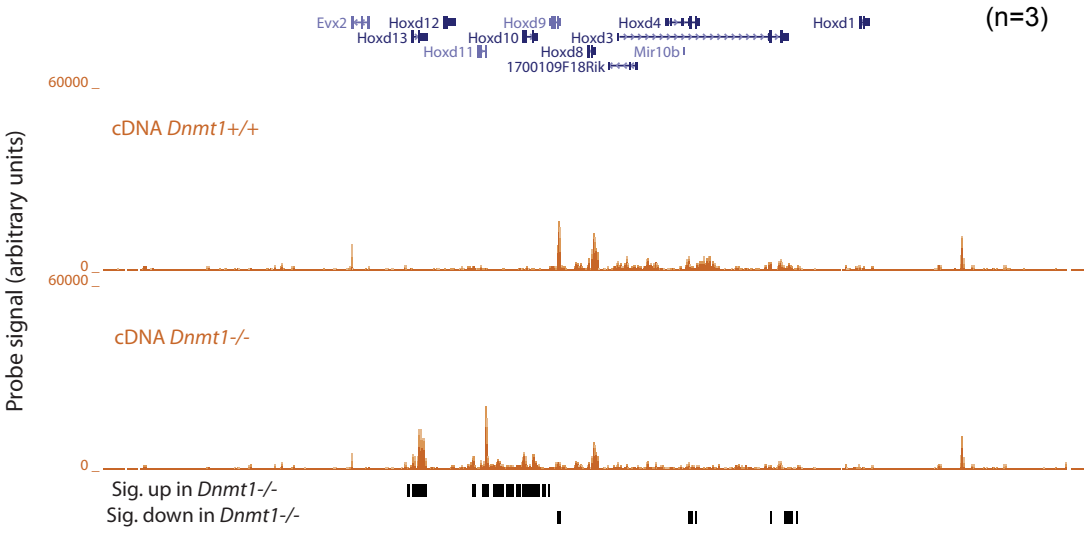



F
Tex19.1
cDNA Dnmt1+/+
cDNA Dnmt1-/-
Probe signal (arbitrary units)
G
Actb
cDNA Dnmt1+/+
cDNA Dnmt1-/-
Probe signal (arbitrary units)

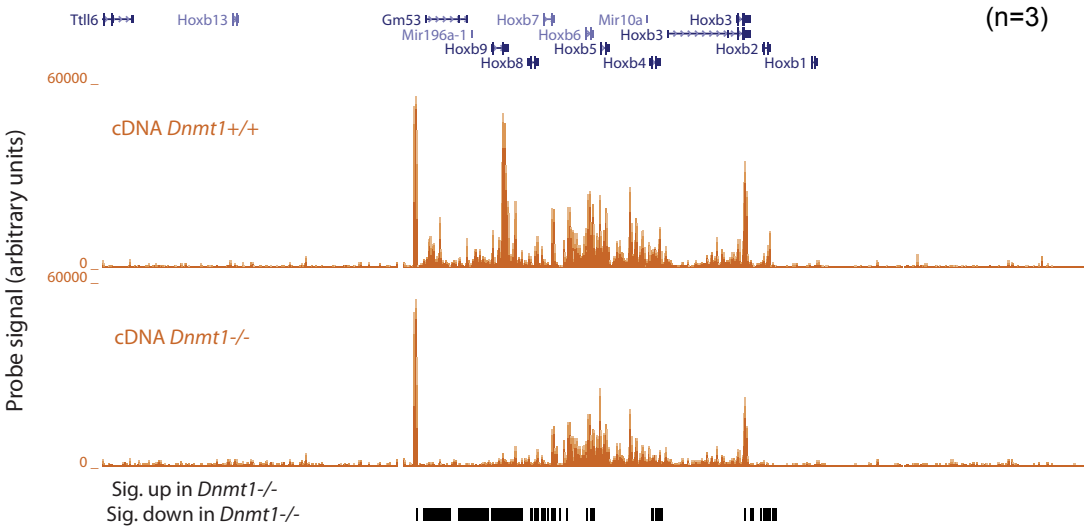

H
(n=3)
Hoxa1
Hoxa2
Hoxa3
Hoxa4
Hoxa5
Hoxa6
2700086A05Rik
Mira
Hoxa7
Hoxa9
Mir196b
Hoxa10
Hoxa10
Hoxa11
Hoxa11as
Hoxa13
60000
cDNA Dnmt1+/+
0
60000
cDNA Dnmt1-/-
0
Probe signal (arbitrary units)
Sig. up in Dnmt1-/-
Sig. down in Dnmt1-/-


I
(n=3)
Hoxc13
Hoxc12
Hoxc11
Hoxc10
Mir196a-2
Hoxc9
Hoxc8
Hoxc6
Hoxc5
Mir615
Hoxc4
60000
cDNA Dnmt1+/+
0
60000
cDNA Dnmt1-/-
0
Probe signal (arbitrary units)
Sig. up in Dnmt1-/-
Sig. down in Dnmt1-/-

J


(n=3)
Evx2
Hoxd12
Hoxd13
Hoxd11
Hoxd9
Hoxd10
Hoxd8
Hoxd4
Hoxd3
Mir10b
1700109F18Rik
Hoxd1
60000
cDNA Dnmt1+/+
0
60000
cDNA Dnmt1-/-
0
Probe signal (arbitrary units)
Sig. up in Dnmt1-/-
Sig. down in Dnmt1-/-

K


(n=3)
Ttll6
Hoxb13
Gm53
Mir196a-1
Hoxb7
Hoxb6
Hoxb9
Hoxb8
Mir10a
Hoxb5
Hoxb4
Hoxb3
Hoxb3
Hoxb2
Hoxb1
60000
cDNA Dnmt1+/+
0
60000
cDNA Dnmt1-/-
0
Probe signal (arbitrary units)
Sig. up in Dnmt1-/-
Sig. down in Dnmt1-/-

decreased expression of many *Hoxa1-6* and increased expression of *Hoxa7-13* genes in *Dnmt1*$^{-/-}$ MEFs by expression microarray (Fig. 3.3C). I noticed that the pattern of *Hox* gene mis-expression observed in *Dnmt1*$^{-/-}$ MEFs was very similar to that observed in mutants in the polycomb gene, *Bmi1*, in MEFs (Cao *et al.*, 2005)(example for *HoxA* cluster, Fig. 3.3C and D). In order to characterise the expression changes at *Hox* clusters in more detail I analysed cDNA from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using a custom Nimblegen tiling microarray containing all four *Hox* clusters represented by densely tiled probes (experimental overview shown in Fig. 3.3E). This microarray will be referred to as the 'custom tiling array' herein to avoid confusion with the other microarray formats used. This array was designed by the lab of Wendy Bickmore to contain densely tiled probes for a total of 556 genomic regions, enriched in PcG target genes and controls, including all four *Hox* clusters (a list of tiled regions covered on this array is provided in Appendix 1, the array format is summarised in the Methods section). Three replicates were performed and probes with significantly different signals were identified (see Methods for details). This experiment confirmed a cluster location effect on relative expression of the *Hox* genes in *Dnmt1*$^{-/-}$ MEFs. At the *HoxA*, *HoxC* and *HoxD* clusters, genes from approximately half of the clusters (~*Hox1-7*) are expressed in *Dnmt1*$^{+/+}$ while genes from the entire clusters are expressed in *Dnmt1*$^{-/-}$ MEFs (Fig. 3.3H to J). Some of the *Hox1-7* genes show decreased expression in *Dnmt1*$^{-/-}$ MEFs at the *HoxA* and *HoxD* clusters, while this was not observed at the *HoxC* cluster where all genes show increased expression in *Dnmt1*$^{-/-}$ MEFs (Fig. 3.3H to J). This analysis also confirmed the striking de-repression of *Hox9-13* genes at the *HoxA*, *HoxC* and *HoxD* clusters in *Dnmt1*$^{-/-}$ MEFs (Fig. 3.3H to J). A similar expression pattern was not observed at the *HoxB* cluster where many genes show decreased expression in *Dnmt1*$^{-/-}$ MEFs and no probes showed significant increased expression (Fig. 3.3K). Overall, the expression changes observed in *Dnmt1*$^{-/-}$ MEFs at the *HoxA*, *HoxC* and *HoxD* clusters are very similar to those reported in *Bmi1*$^{-/-}$ MEFs by Cao *et al.* (2005). As *Hox* genes are known to require PcG proteins for repression, I hypothesised that PcG-mediated repression at *Hox* genes is lost in *Dnmt1*$^{-/-}$ MEFs.

3.2.4. Loss of polycomb-associated K27me3 at HoxC in Dnmt1$^{-/-}$ *MEFs*

As PcG proteins are major regulators of *Hox* gene expression, and given the similarities between *Dnmt1* and *Bmi1* mutation in *Hox* gene mis-regulation, I investigated the presence of the polycomb-associated H3K27me3 and TrxG-associated H3K4me3 (K4me3) marks at the *HoxC* cluster. The K27me3 mark is frequently used as a marker for repression by the

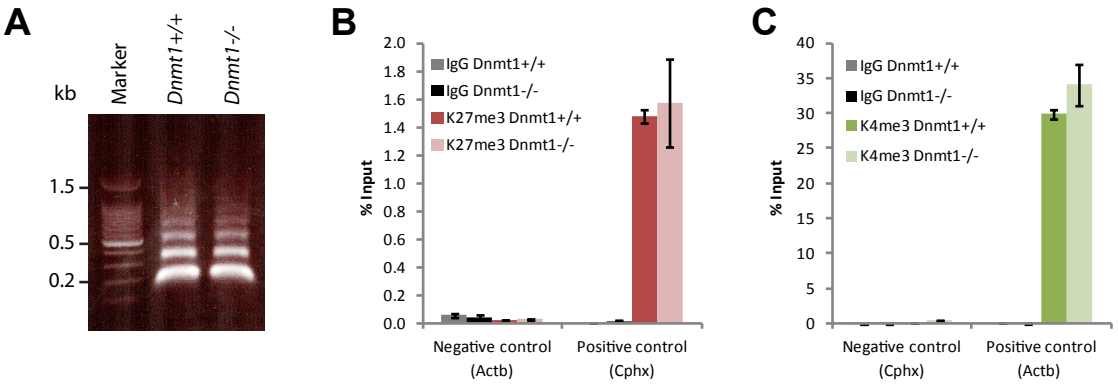
PRC2 complex and K4me3 catalysed by TrxG complexes is thought to be involved in activation of *Hox* genes (Terranova *et al.,* 2006). In order to measure the presence of histone marks, I used native chromatin immunoprecipitation followed by qPCR for primers at the promoter region of each gene (N-ChIP-qPCR). N-ChIP was used as it has certain benefits over techniques involving cross-linked chromatin, such as reduced background and avoidance of potential artefacts introduced by the cross-linking procedure (O'Neill and Turner, 2003). First, I checked that chromatin from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs was digested with similar efficiency by micrococcal nuclease (Mnase) as marked differences here would result in incomparable results. Bulk soluble chromatin from the two cell lines showed no gross differences in digestion by a titration of Mnase (Fig. 3.4A, a single representative digest is shown). Positive and negative control regions for each chromatin mark were tested, along with pull-downs with non-specific IgG, to ensure that N-ChIP was working effectively and background was low (Fig. 3.4B and C). The positive control regions showed similar enrichments between the two cell lines suggesting that N-ChIP from these cell lines are comparable (Fig. 3.4B and C). *Hox* clusters are marked by large 'blocks' of K27me3 in ES cells and MEFs (Mikkelsen *et al.,* 2007; Pauler *et al.,* 2009). At the *HoxC* cluster, as expected, all promoter regions tested showed strong enrichment for K27me3 in *Dnmt1*$^{+/+}$ MEFs. However, a striking loss of K27me3 was observed in *Dnmt1*$^{-/-}$ MEFs that occurred at all promoters tested (Fig. 3.4D). The enrichment for K27me3 at these regions in *Dnmt1*$^{-/-}$ MEFs was comparable to that observed with non-specific IgG (data not shown). This observation is reproducible as many replicates of N-ChIP showed the same result and also using two different antibodies for K27me3 (data not shown). In contrast, a small enrichment over the negative control for K4me3 was observed at *HoxC* promoters in *Dnmt1*$^{+/+}$ MEFs (Fig. 3.4E). In *Dnmt1*$^{-/-}$ MEFs K4me3 was unchanged over *Hoxc4, Hoxc5* and *Hoxc8* promoters but was increased over *Hoxc10* and *Hoxc13,* which also show strong transcriptional derepression in *Dnmt1*$^{-/-}$ MEFs (Fig. 3.4D and Fig. 3.3I). These results suggest that PcG and TrxG regulation of the *HoxC* cluster is altered in *Dnmt1*$^{-/-}$ MEFs.

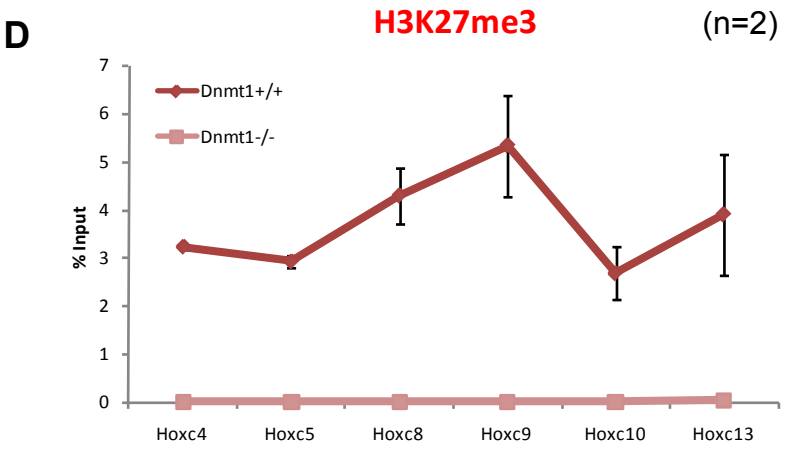
3.2.5. Treatment of cells with demethylating agent 5-aza-dC partially recapitulates the effect of Dnmt1 mutation

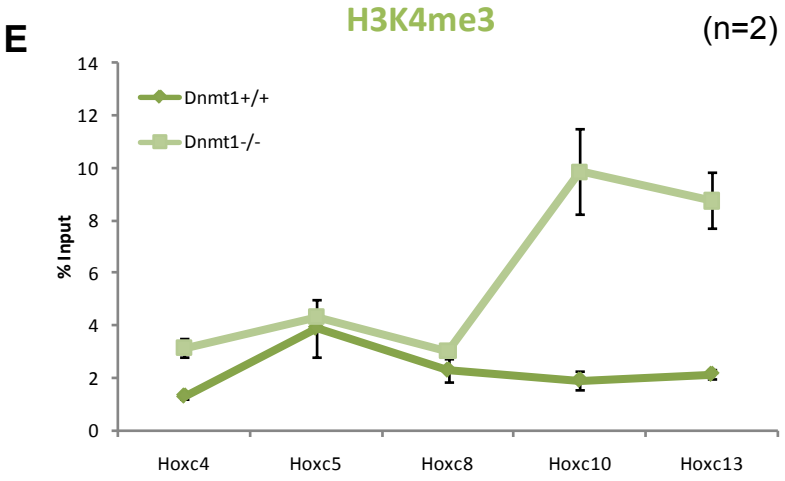
As MEFs are a heterogeneous population of cells I felt that it was crucial to replicate these observations using another system to ensure that they were not artefacts of cell origin. The demethylating agent, 5-aza-dC can be used as an independent method to induce

Figure 3.4 - Loss of polycomb-associated H3K27me3 at *HoxC* genes in *Dnmt1*$^{-/-}$ MEFs

A. Representative agarose gel showing DNA extracted from chromatin preparation digested with miccrococcal nuclease for N-ChIP. B and C. Positive and negative controls for H3K27me3 and H3K4me3 N-ChIP. Error bars represent standard error of the mean. D and E. N-ChIP-qPCR at promoter regions of *HoxC* genes for H3K27me3 and H3K4me3 respectively. Error bars represent standard error of the mean (n=2).







hypomethylation and Dnmt1 depletion (Patel *et al.*, 2010). One of the advantages of 5-aza-dC treatment is that the same cells can be compared with and without the drug, ruling out any effects of caused by heterogeneity of cells. First, *Dnmt1*$^{+/+}$ MEFs were cultured with or without (DMSO control) two different concentrations of 5-aza-dC and gene expression was measured by qRT-PCR. The expression of selected genes that are K27me3 marked and upregulated in *Dnmt1*$^{-/-}$ MEFs, including *HoxC* genes, was measured upon 3 day treatment with 5-aza-dC (Fig. 3.5A). Some, but not all, of these genes showed increased expression with 5-aza-dC treatment at both concentrations of the drug. It should be noted that none of the expression changes are significantly different by two-tailed Student's t-test, probably owing to a large variability in response to the drug and inaccuracy introduced when quantifying low abundance transcripts. Importantly however, treatment with 5-aza-dC resulted in statistically significant reduction of K27me3 at many of the promoters tested while K27me3 enrichment at control genes was either unchanged or increased (Fig. 3.5B). Small increases in K4me3 were observed at some promoters tested but none were found to be statically significant (Fig. 3.5C). To exclude the possibility that these observations are an artefact of *Dnmt1*$^{+/+}$ MEFs, owing for example to their lack of p53, I treated an independent mouse embryonic fibroblast cell line (NIH3T3 cells) with 5-aza-dC. This experiment was more difficult to perform as, at the concentrations of drug used, a higher proportion of cell death was observed. Consequently, 2 day drug treatments were performed, so the results may be expected to show a smaller effect than in *Dnmt1*$^{+/+}$ MEFs. Nevertheless, upon treatment with 5-aza-dC increased expression was observed at the positive control gene *Tex19.1* as well as the *Hox* genes tested (Fig. 3.6A). No statistics were conducted as only one replicate was performed. N-ChIP for K27me3 showed decreased enrichment of the mark at most promoters test but no decrease of a control promoter *Cphx* (Fig. 3.5B). N-ChIP for K4me3 showed an increased enrichment at some *HoxC* promoters but not all upon 5-aza-dC treatment. Also, the changes in K4me3 in these cells did not necessarily correspond to those observed in *Dnmt1*$^{+/+}$ MEFs. Together, these results suggest that the effect of *Dnmt1* mutation on K27me3 can be at least partially recapitulated by treating cells with hypomethylating agent 5-aza-dC. Further validation here is required, particularly for the gene expression assays.

3.2.6. Polycomb target genes are mis-expressed in Dnmt1-mutant embryos

Many polycomb-target genes have important roles in embryonic development (Boyer *et al.*, 2006; Bracken *et al.*, 2006). I chose to investigate the effect of *Dnmt1* mutation on the

Figure 3.5 - Treatment of *Dnmt1*$^{+/+}$ MEFs with demethylating agent 5-aza-dC partially recapitulates the effect of *Dnmt1* mutation at selected PcG target genes

A. qRT-PCR for selected genes, that are up-regulated in *Dnmt1*$^{-/-}$ MEFs, in *Dnmt1*$^{+/+}$ MEFs treated with varying concentrations of 5-aza-dC (AZA). Error bars represent standard error of the mean. Only positive error bars are shown as scale is logarithmic (n=3). None of these changes are significant by two-tailed Student's t-test ($p<0.05$). B and C. N-ChIP-qPCR for H3K27me3 and H3K4me3 resectively in *Dnmt1*$^{+/+}$ MEFs treated with 5-aza-dC. Error bars represent +/- standard error of the mean (n=2). * $p<0.05$ by two-tailed Student's t-test.

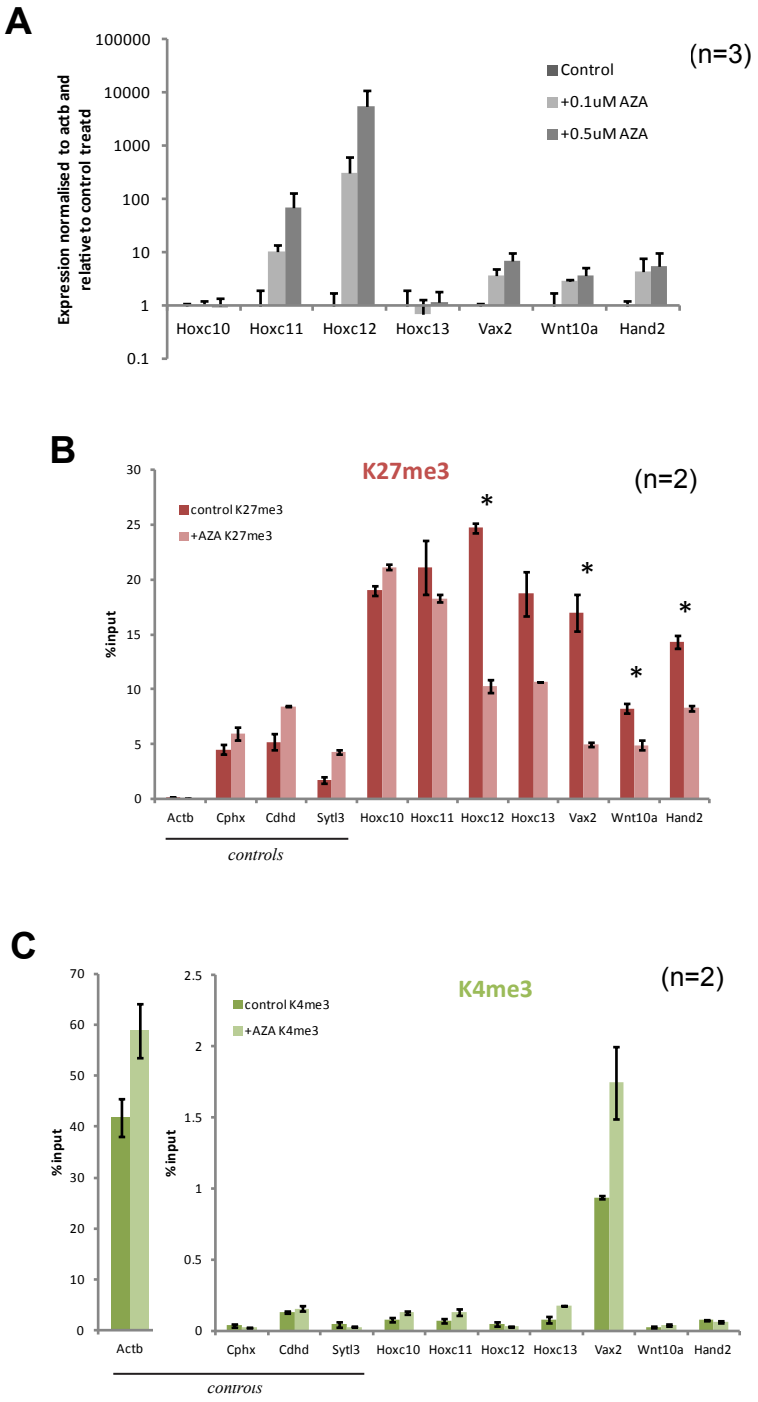
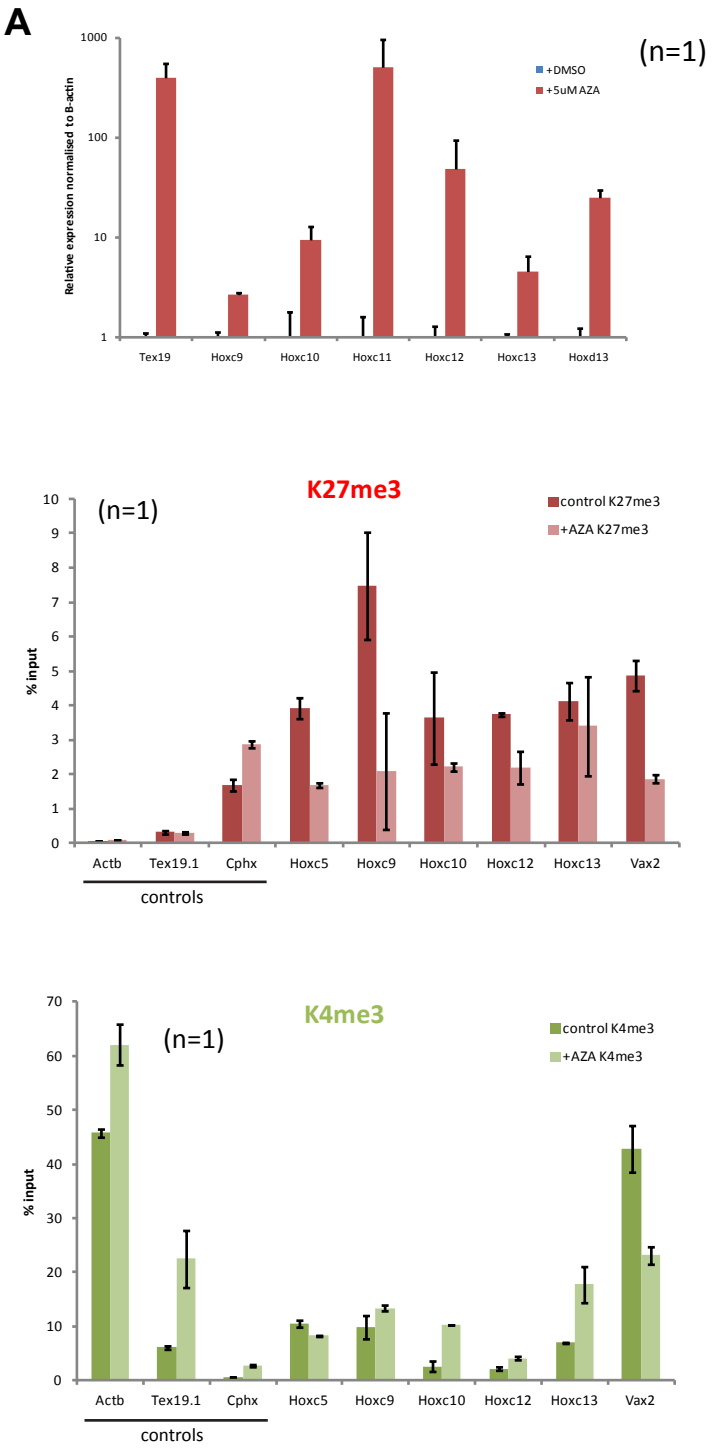


Figure 3.6 - Treatment of NIH3T3 MEFs with demethylating agent 5-aza-dC partially recapitulates the effect of *Dnmt1* mutation at selected *Hox* genes

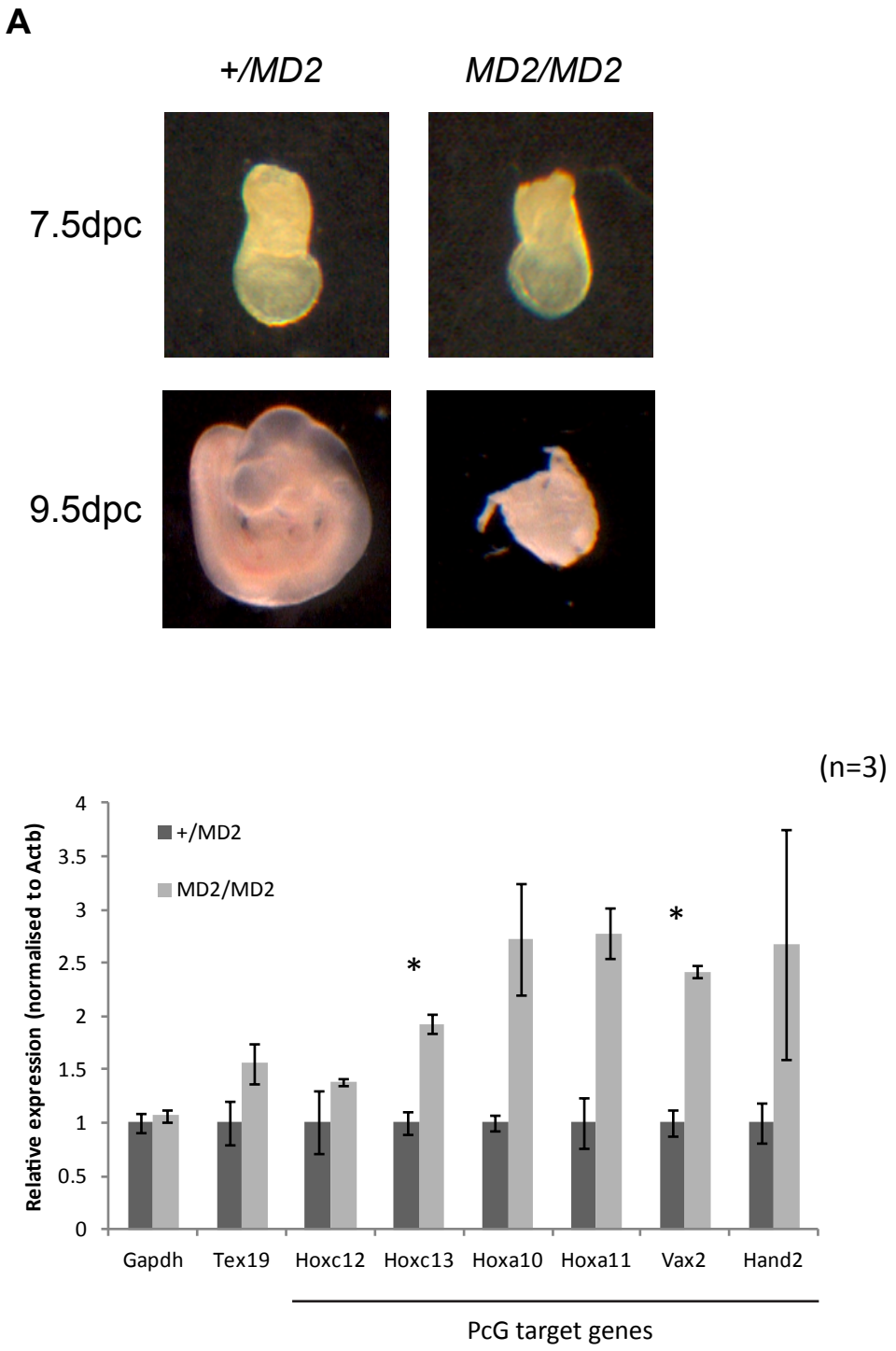
A. qRT-PCR for selected *Hox* genes, that are up-regulated in *Dnmt1*$^{-/-}$ MEFs, in NIH3T3 cells treated with 5-aza-dC (AZA). *Tex19.1* is used as a positive control. Error bars represent standard error of the mean of qPCR technical replicates. Only positive error bars are displayed due to log scale. (n=1). B and C. N-ChIP-qPCR for H3K27me3 and H3K4me3 respectively in NIH3T3 cells treated with 5-aza-dC. Error bars represent standard error of the mean of qPCR technical replicates. (n=1).



expression of PcG target genes in mouse embryos for two reasons; 1) for further confirmation of a role for Dnmt1 in their regulation using a non-cell line system, 2) to investigate if their mis-expression correlates with the developmental phenotype of *Dnmt1*-mutant embryos. Severe mutation in *Dnmt1* results in embryonic lethality and developmental arrest with a failure to properly specify somites along the body axis (Lei *et al,*. 1996). The cause of these phenotypes are unknown but it is compatible with mis-expression of developmental genes, such as *Hox* genes, that are required for these processes. In order to address this we established a collaboration with N.Youngson and E.Whitelaw (QIMR, Brisbane) who had generated mice carrying a point mutation in *Dnmt1* (*MommeD2* allele, or *MD2*)(Chong *et al.*, 2007). The point mutation is within the part of the gene that encodes one of the BAH domains and results in hypomorphic Dnmt1 expression due to protein instability (Chong *et al.*, 2007). This mutation is embryonic lethal when homozygous. Mice heterozygous for *MD2* were crossed and the offspring were dissected, genotyped and imaged by N.Youngson. At 9.5dpc, when other mutations in *Dnmt1* show developmental retardation, *MD2/MD2* embryos showed a similar phenotype and superficially appeared to have failed to specify a proper body axis (Fig. 3.7A). +/*MD2* embryos were indistinguishable from +/+ at this stage. As gene expression in *MD2/MD2* and +/*MD2* embryos at this stage would be incomparable due to the extreme morphological differences caused by developmental defects in the *MD2/MD2* embryos, we looked earlier in development, at 7.5dpc. At this stage, +/*MD2* and *MD2/MD2* embryos were indistinguishable at by gross morphology (Fig. 3.7A). Therefore, we decided to investigate the expression of PcG target genes at this stage, before the developmental phenotype became apparent (at least at this resolution). Using three +/*MD2* and three *MD2/MD2* embryos I analysed gene expression from total embryo RNA that was extracted by N.Youngson. All but one of the PcG target genes shown in figure 3.7B showed increased mean expression in *MD2/MD2* relative to +/*MD2* embryos but only two were statistically significant by Student's t-test. The expression of these genes showed considerable variability between individuals leading to lack of statistical significance. This probably arises through both variation in dissection of embryos at this early stage and also through temporal developmental variation between individual embryos. Analysis of more embryos may allow significance to be reached. The expression of *Tex19.1* was also examined as it was suspected to be a positive control for hypomethylation. Surprisingly this gene did not show strong upregulation in *MD2/MD2* embryos. In retrospect, this is likely due to the fact that *Tex19.1* is normally expressed in some embryonic and extra-embryonic tissues at this stage and so is not strongly de-repressed (Hackett, 2010, PhD thesis; Hackett *et al.*, in submission).

Figure 3.7 - Polycomb target genes are mis-expressed in *Dnmt1*-mutant embryos

A. Representative pictures of embryos at 7.5dpc and 9.5dpc that are heterozygous or homozygous for the *MommeD2* point mutation in *Dnmt1* (*MD2*). (Performed by N. Youngson, QIMR). Mutation described in Chong et al. (2007). B. qRT-PCR showing relative expression of the indicated polycomb target genes in 7.5dpc embryos that are heterozygous or homozygous for the *MD2* point mutation. Error bars represent the standard error of the mean (n=3). * $p<0.05$ by Student's t-test.



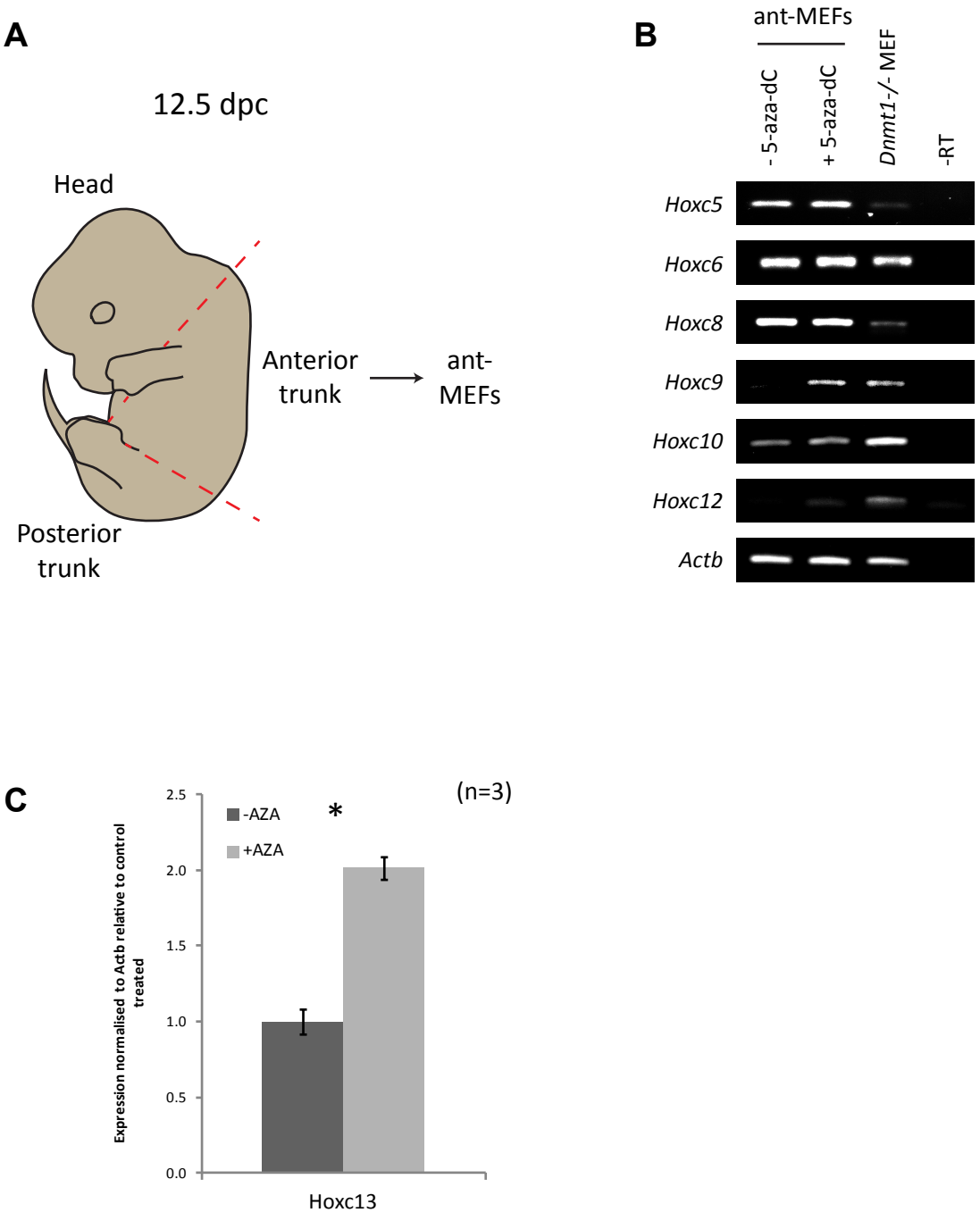
These data provide suggestive evidence that Dnmt1 may be required in the early mouse embryo for repression of key developmental genes that are known PcG targets.

3.2.7. Derivation of 'anterior trunk' MEFs and treatment with 5-aza-dC

Analysis of *Hox* gene mis-expression in *Dnmt1*$^{-/-}$ MEFs showed that *Hox9-13* genes from three of the *Hox* clusters are strongly de-repressed (Fig. 3.3). These genes are normally expressed in the very posterior portion of the mouse embryo in a nested fashion dependent on their position within the cluster (Peterson *et al.*, 1994; Nordemeer *et al.*, 2011). MEFs are derived from decapitated, but otherwise whole, embryos and so the vast majority of cells would be expected to represent anterior regions as the *Hox9-13* expressing posterior regions represent a small proportion of the embryo. This suggests that Dnmt1 and/or DNA methylation may be required to maintain repression of these genes in anterior portions of the embryo. To test this hypothesis I decided to investigate the effect of 5-aza-dC treatment on MEFs derived from the anterior trunk portion of the embryo. I dissected 12.5 dpc wildtype CD1 embryos from their deciduas and extra-embryonic tissue, decapitated them and then removed the posterior trunk by cutting above the hind limb buds (Fig. 3.8A). Using the remaining anterior trunk segment I derived MEFs (anterior trunk MEFs, or, ant-MEFs). I performed RT-PCR to check for expression of *HoxC* genes in ant-MEFs (Fig. 3.8B first column). As expected these cells express high levels of the more anterior-expressed *Hoxc5*, *Hoxc6* and *Hoxc8* and low levels of the more posterior-expressed *Hoxc9*, *Hoxc10* and *Hoxc12* genes suggesting that they maintain their positional identity in culture. Next, I treated ant-MEFs with 1µM 5-aza-dC for three days. This higher dose of 5-aza-dC was applied as I expected these primary MEFs to divide much slower, and hence incorporate less of the drug, than the immortal fibroblasts used in the previous treatments with 5-aza-dC. 5-aza-dC treatment resulted in increased expression of *Hoxc9* and *Hoxc12* by RT-PCR (Fig. 3.8B). As the most posterior-expressed *HoxC* gene, *Hoxc13* relative expression was measured by qRT-PCR in triplicate. A statistically significant two-fold increase in *Hoxc13* was observed in ant-MEFs treated with 5-aza-dC (Fig. 3.8C, $p<0.05$ by two-tailed Student's t-test). This suggests that Dnmt1 and/or DNA methylation may contribute to the repression of posterior-expressed *HoxC* genes in anterior tissues of the developing mouse embryo.

Figure 3.8 - Treatment of 'anterior trunk' MEFs with 5-aza-dC

A. Representation of a 12.5 dpc mouse embryo indicating the portion of the embryo used to derive 'anterior trunk' mouse embryonic fibroblasts (ant-MEFs). B. RT-PCR for *HoxC* genes in ant-MEFs with and without treatment with 5-aza-dC. *Dnmt1*$^{-/-}$ cDNA used as a positive control. -RT is a no reverse transcriptase control. C. qRT-PCR for *Hoxc13* in ant-MEFs with and without 5-aza-dC treatment for three days. Error bars indicate +/- standard error of the mean (n=3). * $p<0.05$ by two-tailed Student's t-test.



3.3. Discussion

In this chapter, I aimed to gain further insight into the mechanisms by which DNA methylation affects gene repression. I have shown that normal levels of Dnmt1 are required for the repression of certain PcG target genes in mouse embryonic fibroblasts and possibly embryos. Initially, I integrated data for promoter K27me3 and DNA methylation, and gene expression in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs. The caveat of this analysis is that these datasets were generated by different labs using 'MEF' cells. MEFs are a heterogeneous cell population and would be expected to comprise different cell populations depending on the embryonic stage used for derivation and hence may not be directly comparable. For this reason, the analyses in section 3.2.1 were intended only to provide clues which would then be followed up experimentally. Later in this thesis I describe K27me3 and DNA methylation mapping in *Dnmt1*$^{+/+}$ MEFs, addressing this problem.

I focussed my subsequent observations on the *Hox* genes and showed that many of them are mis-expressed in *Dnmt1*$^{-/-}$ MEFs, cells treated with hypomethylating agent 5-aza-dC and in *Dnmt1*-mutant embryos. Importantly, *Hox* mis-expression was observed in multiple experimental contexts, supporting a *bona fide* role for Dnmt1 in the regulation of these genes. This was particularly important as using the MEFs alone it would have been entirely feasible that gene expression differences could be attributed to different tissues of origin or sub-populations of cells within the culture. The fact that treatment with 5-aza-dC causes similar changes to *Dnmt1* mutation also tells us something about the mechanism involved. When studying gene expression and chromatin patterns in mutant MEFs we are studying an endpoint and it is difficult to infer when the changes have arisen or what has caused them. *Dnmt1*$^{+/+}$ MEFs have an established chromatin and expression state at *Hox* genes but these can be altered by 5-aza-dC treatment supporting a direct role for Dnmt1 in this process. In respect to expression of the *Hox* genes studied, the *Dnmt1*$^{-/-}$ MEF cell line used here appears to show similar effects to other experimentally hypomethylated cell lines and embryos, making them a valuable tool to study the function of Dnmt1 and DNA methylation.

The observations presented here link DNA methylation to *Hox* gene regulation and other PcG target genes but this is not the first time this connection has been made. Mutations of *Dnmt3b* that are associated with ICF syndrome in humans lead to a homeotic phenotype when expressed in mice accompanied by mis-expression of *Hox* genes (Velasco *et al.*, 2010). Nearly half of genes mis-expressed in cells from ICF patients with DNMT3B

mutations are normally PcG targets and include key developmental genes (Jin *et al.*, 2008; Jin *et al.,* 2009). It could be conceived that mild mis-expression of some of these genes could contribute to the phenotypes of ICF syndrome, which include developmental abnormalities of cranio-facial development. Also, mutation of *DNMT3A* is frequent in patients of certain leukaemias, where decreased DNA methylation and increased expression of some *Hox* genes has been observed (Yan *et al.,* 2011). Mutation of *Lsh*, a gene encoding a supposed DNA methylation co-factor, leads to decreased DNA methylation and mis-expression of certain *Hox* genes in cell lines and tissues (Xi *et al.,* 2007; Tao *et al.,* 2010). It should be noted however, that as the DNMTs have reported functions out with of DNA methylation, it cannot be formally excluded at this stage that these non-catalytic roles are important here. It seems likely however, that the DNA methylation mark is involved, as *Hox* gene mis-expression is observed in the multiple contexts described above where decreased DNA methylation is observed. Despite these observations, the cause of *Hox* and other PcG target gene mis-expression in mutants of the DNA methylation machinery is not understood. The results presented here are novel amongst previously published work for two reasons; 1) *Hox* mis-expression has not yet been reported in the detail included here or in such a severely hypomethylated cell system as the *Dnmt1*$^{-/-}$ MEFs, and, 2) striking losses in polycomb-associated K27me3 have not been reported in hypomethylated cells. This second point is important as it offers a potential explanation for *Hox* mis-expression when cells are hypomethylated. At this point it is not clear whether K27me3 loss is upstream and the cause of *Hox* mis-expression or if it is merely a consequence of transcriptional upregulation. It is clear from the literature however, that loss of PcG binding at *Hox* genes causes dramatic mis-expression (Cao and Zhang, 2004). The cause and consequence relationships here remain to be tested but the results presented in this chapter suggest a further link between DNA methylation and *Hox* gene regulation.

Hox genes were observed to be upregulated in mouse embryos depleted in Dnmt1 before the onset of gross morphological phenotypes. This is an interesting suggestion as the cause of the developmental phenotypes in *Dnmt1*-mutants is not clear. At least part of the phenotype has been attributed to increased apoptosis in the developing embryo (Jackson-Grusby *et al.*, 2001). Developmental phenotypes such as failure to specify somites and incomplete closure of the neural tube are also observed (Lei *et al.*, 1996). One possibility for these phenotypes is mis-expression of developmental genes, although to date this has not been studied directly in the embryo to my knowledge. It is interesting to speculate that mis-expression of developmental genes, such as *Hox* genes, contributes to these phenotypes as they have important roles in these processes (Soshnikova and Duboule, 2009a). In this

regard, the experiment shown in figure 3.8 suggests that Dnmt1 may be required for repression of normally posterior-expressed *HoxC* genes in anterior tissues. The embryo observations presented in this chapter are preliminary for a number of reasons. Firstly, *in situ* hybridisation would need to be performed on *Dnmt1*-mutant embryos to better characterise mis-expression of *Hox* genes. Secondly, a much more thorough analysis of the phenotype is required to make sure that the structures of the mutant embryo at 7.5dpc are comparable. It would be interesting to examine *Hox* gene mis-expression at multiple time points during early embryogenesis and correlate this with the emergence of certain body axis phenotypes. This kind of study may be more easily achieved in another vertebrate model organism such as zebrafish or *Xenopus laevis* where somites and body axis phenotypes are more easily visible. This would also address the question of evolutionary conservation of this mechanism among vertebrates. Interestingly in this regard, treatment of zebrafish embryos with 5-aza-dC and depletion of xDnmt1 in *Xenopus laevis* reportedly result in body plan and somite patterning defects (Martin *et al.*, 1999; Stancheva and Meehan, 2000). Further characterisation of ant-MEF response to 5-aza-dC could prove informative, including N-ChIP to investigate K27me3 levels upon treatment. N-ChIP of *Dnmt1*-mutant embryos would also be informative to connect de-repression to K27me3 changes. For these experiments a conditional knock-out allele for *Dnmt1* may provide a cleaner experimental approach.

The effects of hypomethylation on K27me3 and K4me3 described here at *Hox* genes are intriguing. At this point, it is not clear if these changes will be specific to *Hox* clusters, a small number of genes, or if they represent a more widespread effect of hypomethylation. As a large number of PcG targets were found to be upregulated in *Dnmt1*$^{-/-}$ MEFs by expression microarray, and upon 5-aza-dC treatment, this remains a possibility. It is interesting to speculate that a major role of DNA methylation may be in shaping these two histone marks, a question that will be addressed in the following chapter.

Chapter 4 - Mapping changes in histone marks in DNA hypomethylated cells

4.1. Introduction

DNA methylation has been strongly linked with patterns of histone modification and has been suggested to contribute to the maintenance of certain chromatin states. Understanding the interplay between epigenetic marks is a big problem in the field and is key to understanding their role in gene regulation. One histone mark that has been linked to DNA methylation is H3K4me3 (K4me3), a modification associated with gene promoter regions and unmethylated CpG islands (Mikkelsen *et al.*, 2007; Barski *et al.*, 2007; Thomson *et al.*, 2010). DNA methylation at CpG-rich promoter regions has been observed to be negatively correlated with K4me3 (Meissner *et al.*, 2008). A candidate for a protein that links DNA methylation and K4me3 is Cfp1, which has been suggested to bind to unmethylated CpG motifs through its CXXC domain and recruit the K4 methylase Set1 complex (Thomson *et al.*, 2010). The role of DNA methylation in shaping K4me3 patterns is yet to be fully addressed. A second histone modification that has been linked to DNA methylation patterns is H3K27me3 (K27me3), the mark deposited by the PRC2 complex. This mark has also been negatively correlated with DNA methylation in a number of studies (see section 1.3.6). The underlying mechanisms for these observations are not understood and the interplay between these two epigenetic layers remains to be fully characterised. In particular, many of the studies linking DNA methylation to these histone modifications have done so by correlation and few direct tests have been performed, particularly for K27me3.

The results presented in chapter 3 demonstrate changes in the histone marks K27me3 and K4me3 at *HoxC* gene promoters in severely hypomethylated *Dnmt1*$^{-/-}$ MEFs. One possible explanation for these results is that DNA methylation is required for correct targeting of these histone marks to these regions. Chapter 3 investigates candidate genes and therefore does not address the possibility that K27me3 and K4me3 patterns are more widely affected in hypomethylated cells. In this chapter I address this problem by mapping K27me3 and K4me3 at gene promoter regions on a large scale in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Promoter DNA methylation mapping in *Dnmt1*$^{+/+}$ MEFs is also performed. By integrating this data with expression microarray findings I study the relationships between gene expression, DNA methylation and histone modification changes in DNA hypomethylated cells. These results suggest a major role for an intact DNA methylation system in shaping the patterns of these key modifications of chromatin, and in turn, their modulation of gene expression.

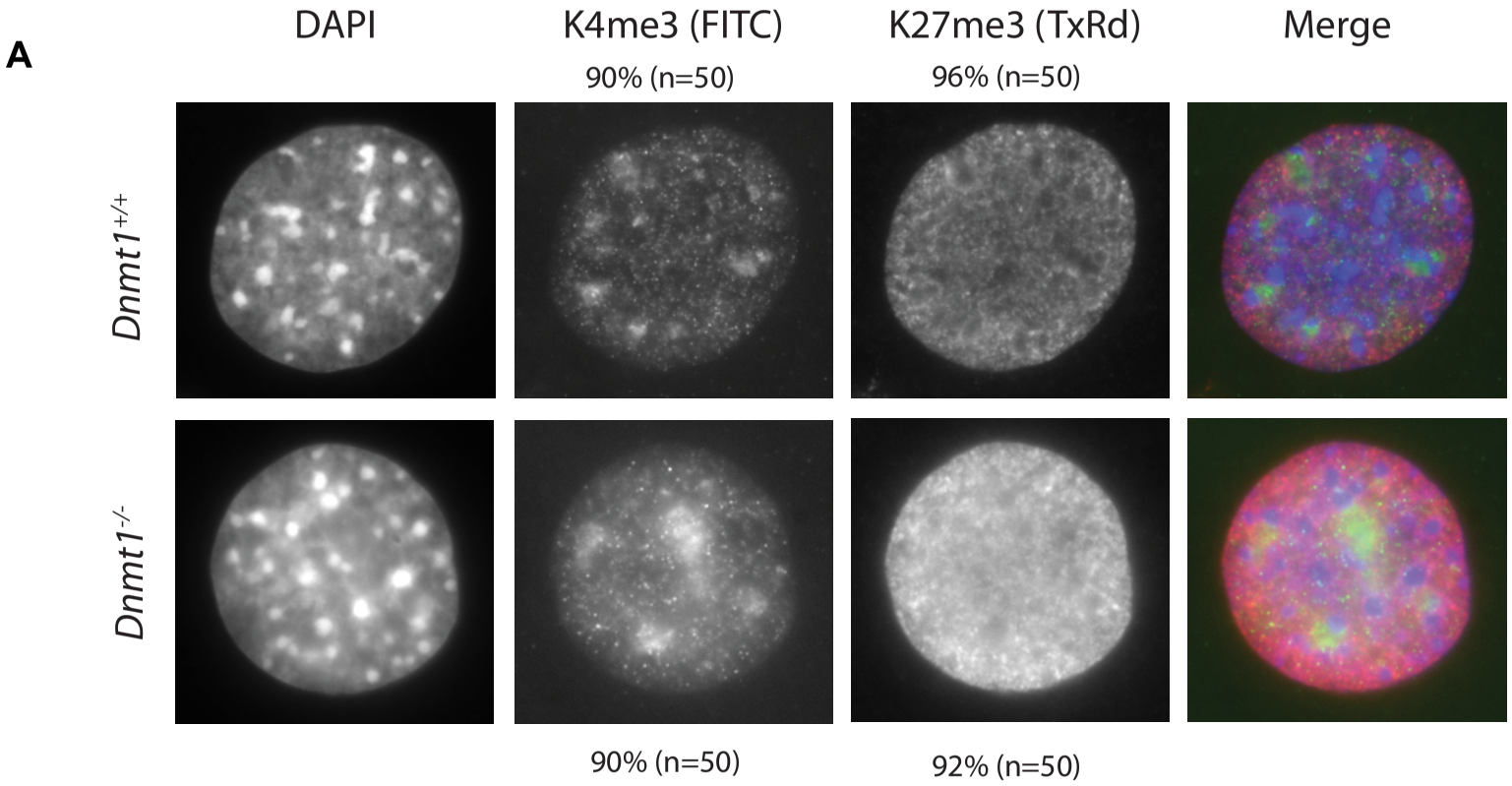
4.2. Results

4.2.1. Global abundance of histone marks and PcG proteins

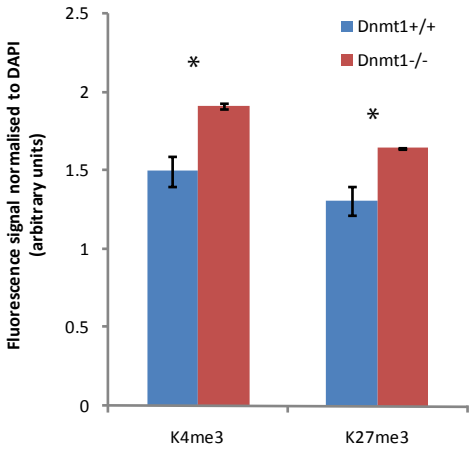
Before mapping histone marks in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs I analysed their global abundance. This is important as changes in histone marks, such as the loss of K27me3 observed at the *HoxC* cluster in *Dnmt1*$^{-/-}$ MEFs (chapter 3), could be attributable to a global decrease in K27me3 levels. In order to assess the global abundance of histone marks I performed immunofluorescence (IF) on cells and western blot analyses of nuclear lysates. DAPI staining of DNA in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs showed the expected 'bright-spots', or chromocenters, that are characteristic of mouse cells and represent mainly densely stained pericentric heterochromatin regions (Fig. 4.1A). The chromocenters were not examined in detail but no drastic changes were observed in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.1A). K4me3 and K27me3 signals were exclusively nuclear and both formed foci, although this was more apparent for K4me3 under these conditions. The observation that was consistent among cells was that both K4me3 and K27me3 signal was higher in *Dnmt1*$^{-/-}$ than *Dnmt1*$^{+/+}$ MEFs (Fig. 4.1A, true for >90% of cells). This increased signal was apparent as an increased general nuclear signal rather than a change in sub-nuclear staining. Changes in the number or size of the larger K4me3 foci, as suggested by the image in figure 4.1A, were not a consistent observation. The fluorescence signal was calculated from fields of cells and normalised to DAPI signal in order to quantify this increase. Both K4me3 and K27me3 showed a modestly increased signal in *Dnmt1*$^{-/-}$ MEFs using this method (Fig. 4.1B). A western blot of histone marks from soluble chromatin from nuclear lysate also suggested increased levels of K4me3 and K27me3 in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.1C). Western blot for H3K4me2 showed no change while H4K20me3 was decreased in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.1C). H2AK119ub produced a very weak signal that was difficult to compare against background (Fig. 4.1C). Next, I performed western blot for PRC2 components in whole cell lysates from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs which showed no major changes in levels of these proteins (Fig. 4.1D). IF using antibodies against the PRC2 component Ezh2 and the PRC1 component Bmi1 showed no major changes in *Dnmt1*$^{-/-}$ MEFs, with both proteins showing a similar speckled localisation (Fig. 4.1E). Together, these results suggest that the loss of K27me3 from the *HoxC* genes is not due to global loss of K27me3 or changes in PRC2 core protein levels. They also suggest that overall levels of K27me3 and K4me3 are slightly increased in *Dnmt1*$^{-/-}$ MEFs.

Figure 4.1 - Global abundance of histone marks and polycomb proteins in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs

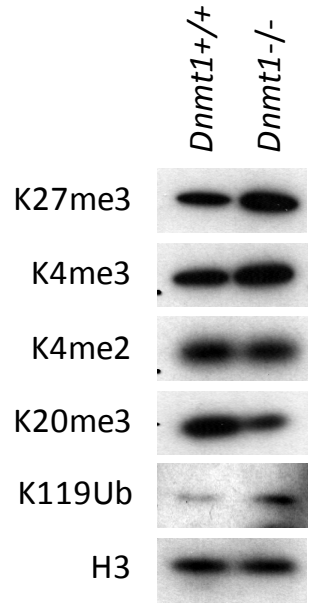
A. Immunofluorescence on *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using antibodies against H3K4me3 and H3K27me3. Representative nuclei shown. Percentage indicates the proportion of nuclei that show the pattern shown in the image. B. Quantification of fluorescence signal from immunofluorescence of *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using antibodies against H3K4me3 and H3K27me3. Signal is normalised to DAPI signal. Error bars represent standard error of the mean. * P<0.05 Student's T-test. C. Western blot on nuclear extract from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using antibodies against various histone modifications. Histone H3 (H3) is used a s a loading control. D. Western blot on whole cell extract from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using antibodies against Ezh2 and Bmi. Actb is used as a loading control. E. Immunofluorescence on *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs using antibodies against polycomb proteins Bmi1 and Ezh2.



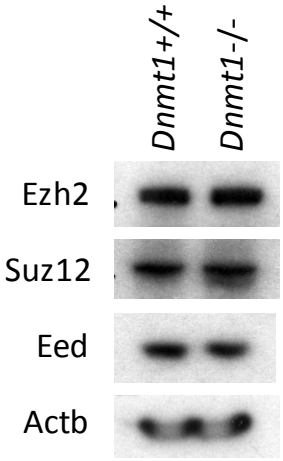
B


Fluorescence signal normalised to DAPI (arbitrary units)
2.5
2
1.5
1
0.5
0
Dnmt1+/+
Dnmt1-/-
*
*
K4me3
K27me3

C


Dnmt1+/+
Dnmt1-/-
K27me3
K4me3
K4me2
K20me3
K119Ub
H3

D


Dnmt1+/+
Dnmt1-/-
Ezh2
Suz12
Eed
Actb

E
Dnmt1-/-
Dnmt1+/+
DAPI
Bmi1 (FITC)
Merge
Bmi1
DAPI
Ezh2 (FITC)
Merge
Ezh2

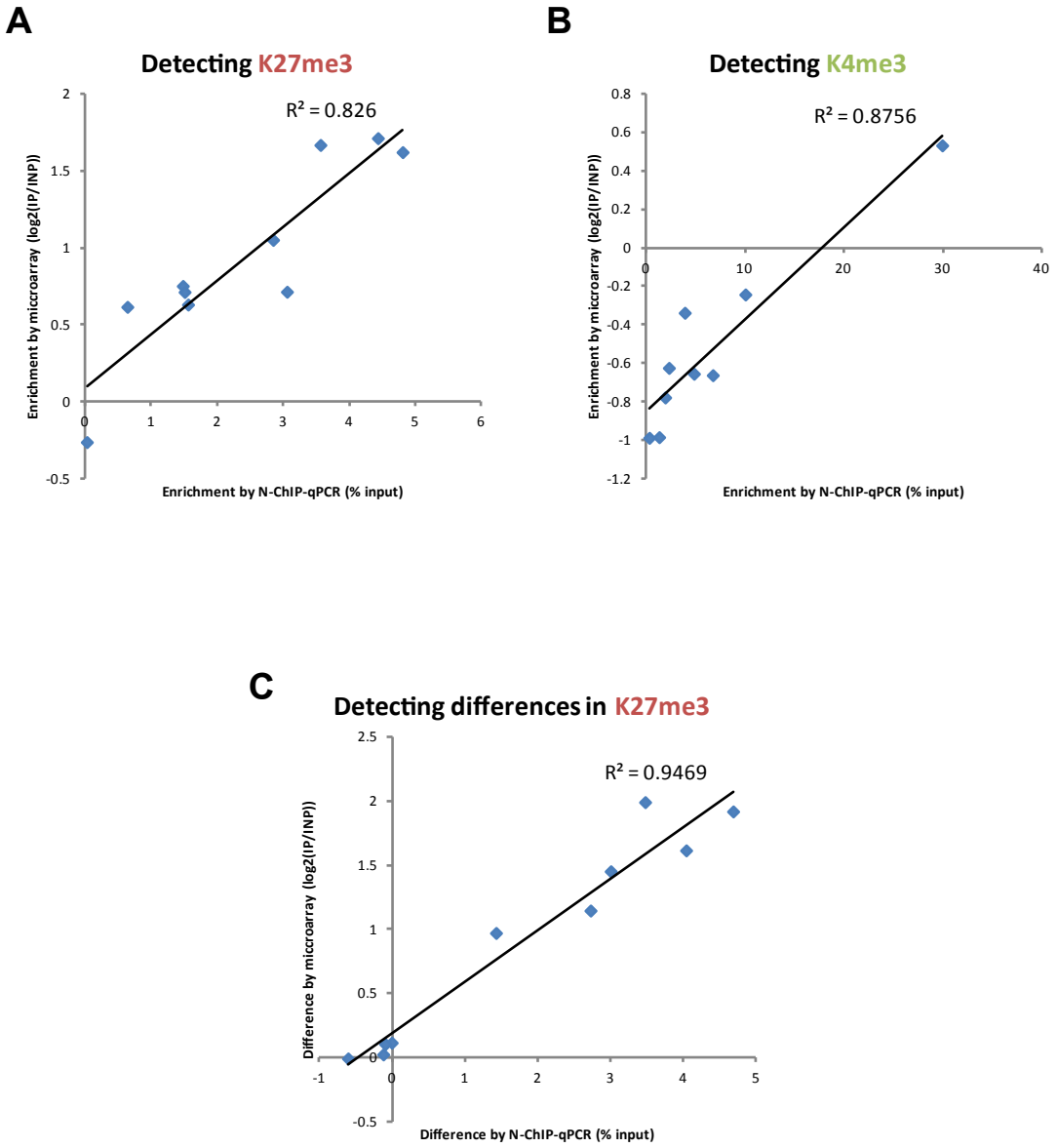
4.2.2. Changes in histone marks occur at a large number of developmental gene promoters in Dnmt1$^{-/-}$ MEFs

Loss of K27me3 at *HoxC* genes (chapter 3) appears to be specific to certain loci as no difference was observed at a control region by N-ChIP-qPCR (chapter 3), and globally by western blot and IF. In order to investigate the changes in K27me3 and K4me3 at gene promoters on a larger scale I performed N-ChIP coupled with microarray (N-ChIP-chip) analysis using a microarray covering a large number of promoter regions. This array is available commercially from Nimblegen (details of array format in Methods) and contains about 20,000 RefSeq promoter regions tiled with a median probe spacing of 100bp. Herein, this microarray format will be referred to as the 'promoter array'. Normalisation and data processing for N-ChIP-chip is described in chapter 2. In order to check that this technique can be used to accurately assess enrichment of these histone marks at specific regions, I picked representative promoters that show a range of enrichments by microarray and validated them by N-ChIP-qPCR. The promoter array produced results very similar to those of N-ChIP-qPCR for detection of K27me3 and K4me3 enrichments in *Dnmt1*$^{+/+}$ MEFs, as assessed by Pearson's correlation (R^2=0.826 and 0.876 respectively)(Fig. 4.2A and B). The accuracy of the microarray technique in detecting differences in enrichment between the *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEF cells was also assessed. The results matched those of N-ChIP-qPCR well for K27me3 differences (R^2=0.947)(Fig. 4.2C). These validations suggest that the microarray technique used, including DNA amplification steps, can be used to accurately map enrichments and cell line differences in histone marks.

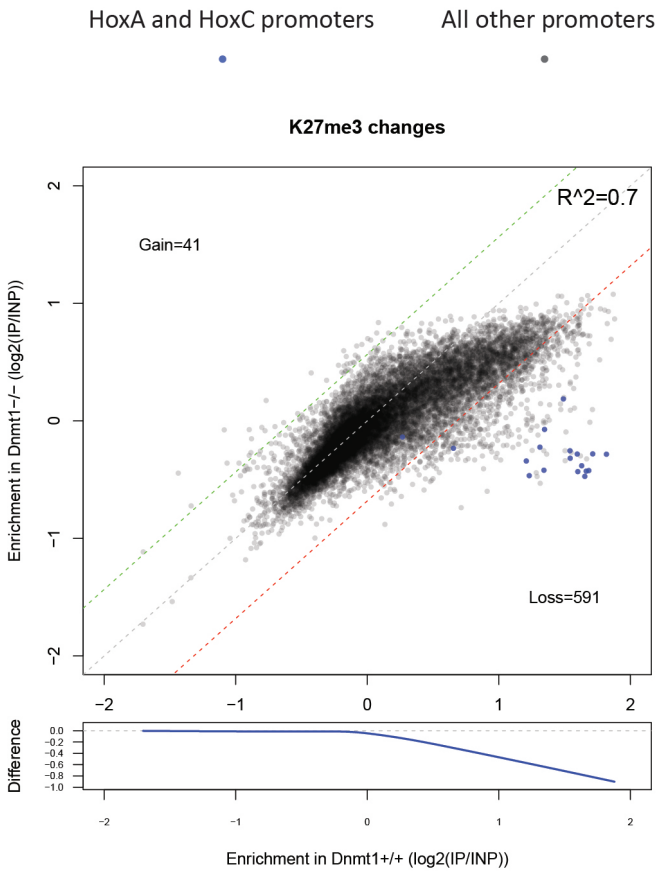
A large number of promoters showed differential K27me3 or K4me3 in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.2D and E upper). Thresholds were set in order to define lists of promoters that were most strongly changed by calculating the difference between the cell lines for each promoter region then selecting difference values that varied from the median difference value by more than four median absolute deviations (see chapter 2 for details)(upper and lower thresholds are shown as dashed green and red lines respectively). These thresholds were used to define 'gain' and 'loss' promoters described herein and refer to signal in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs. This stringent method provided a list of 41 promoters with increased K27me3 and, a surprisingly large, 591 that showed decreased K27me3 in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.2D upper). For K4me3, 294 promoters showed an increase and 41 showed a decrease in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.2E upper). A full list of genes in each category is

Figure 4.2 - Mapping histone modifications at promoters by N-ChIP-chip

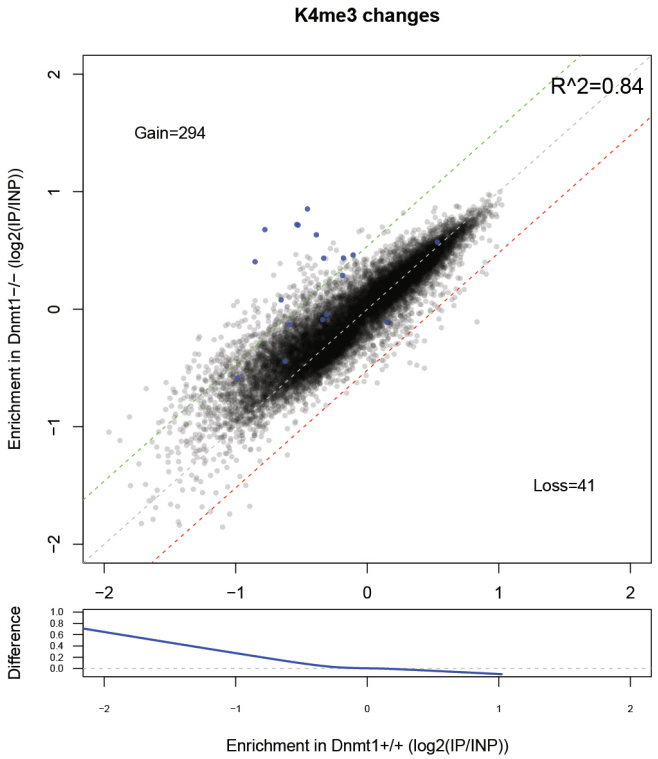
A and B. Scatter plots showing the correlation between N-ChIP-chip and N-ChIP-qPCR for detecting K27me3 and K4me3 enrichments at gene promoters. C. Scatter plot showing correlation between N-ChIP-chip and N-ChIP-qPCR for detecting differences in K27me3 at gene promoters between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. D and E. (Above) Scatter plots of K27me3 and K4me3 signal respectively of all arrayed promoters in *Dnmt1*+/+ and *Dnmt1*-/- as measured by ChIP-chip on Nimblegen promoter array. Red and green dashed lines indicate lower and upper thresholds for change respectively (see methods). Numbers of gene promoters that exceed these thresholds are given as 'gain' or 'loss' in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$. Genes in the *HoxA* and *HoxC* clusters are plotted in blue. (Below) Difference for each promoter region between *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ MEFs as a function of enrichment in



D


HoxA and HoxC promoters
All other promoters
K27me3 changes
R^2=0.7
Gain=41
Loss=591
Enrichment in Dnmt1−/− (log2(IP/INP))
Difference
Enrichment in Dnmt1+/+ (log2(IP/INP))

E

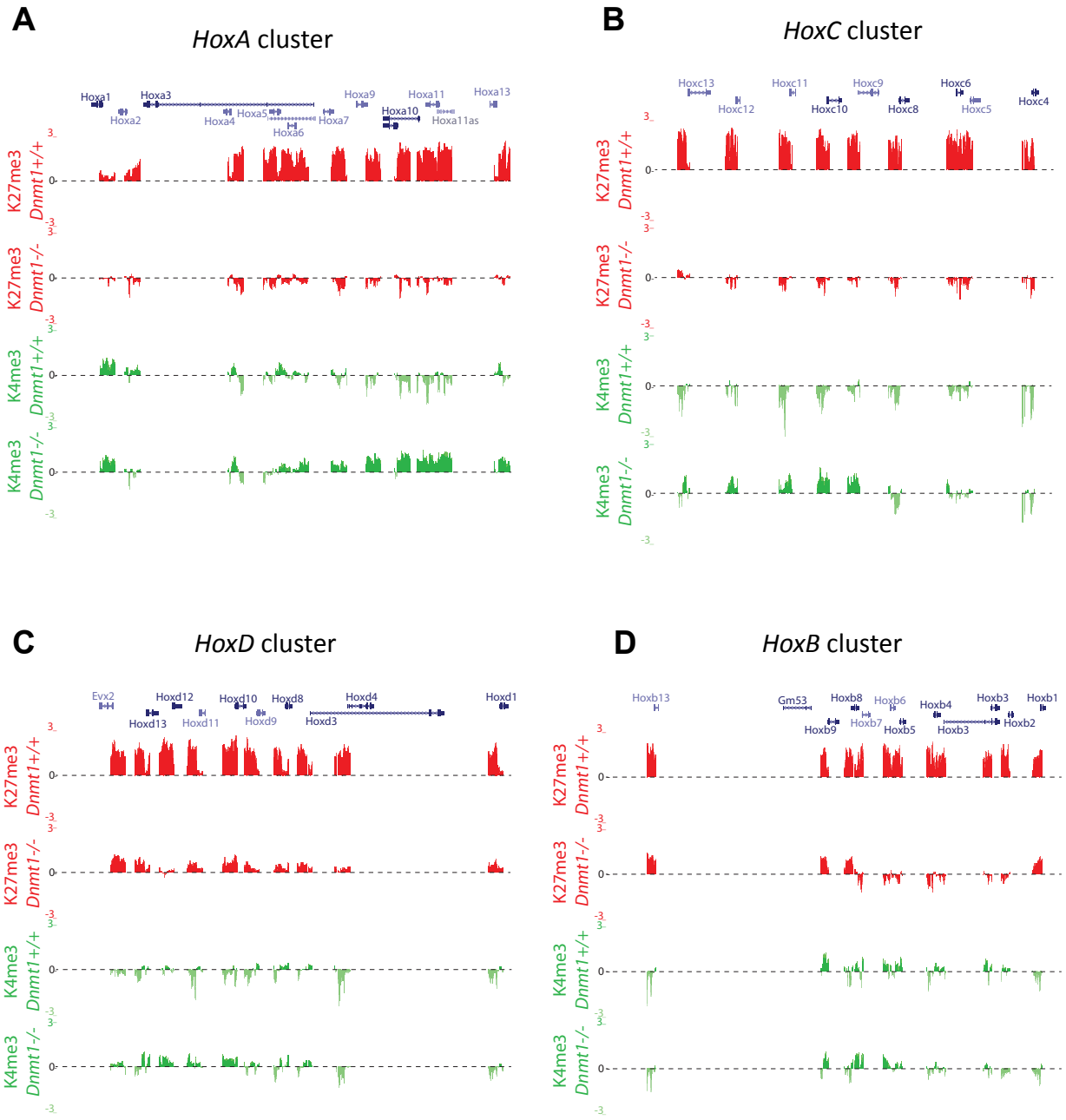

K4me3 changes
R^2=0.84
Gain=294
Loss=41
Enrichment in Dnmt1−/− (log2(IP/INP))
Difference
Enrichment in Dnmt1+/+ (log2(IP/INP))

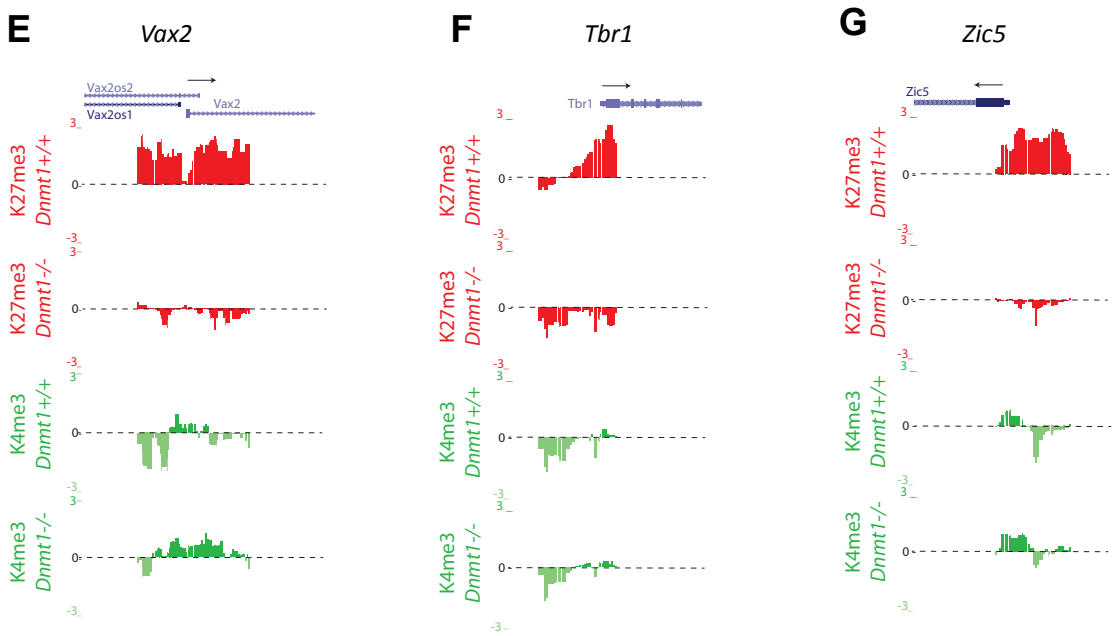
provided in appendix 2. Plotting the points representing promoters of genes in the *HoxA* and *HoxC* clusters in blue revealed that these promoters are among the most changed for both histone marks, showing striking loss of K27me3 and gain of K4me3 in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.2D and E upper). Overall, the K27me3 mark showed a greater variation at promoters in *Dnmt1*$^{-/-}$ MEFs than the K4me3 mark as assessed by their R^2 values (0.7 and 0.84 respectively). What is clear for K27me3 is that promoters with the highest enrichment in *Dnmt1*$^{+/+}$ MEFs show the largest variation and a tendency for K27me3 loss, as indicated by the decreased dependence and increased skewing at higher values of x (Fig. 4.2D upper). This is supported by plotting the difference between the two cell lines as a function of the signal in *Dnmt1*$^{+/+}$ which shows a negative effect of *Dnmt1*$^{+/+}$ signal on difference at positive values of x (Fig. 4.2D lower). These observations suggest that promoters normally highly bound by PcG, require an intact DNA methylation system for full PcG binding. The large number of promoters that lose K27me3 in *Dnmt1*$^{-/-}$ MEFs suggests that this effect is widespread. For the K4me3 mark, it is clear that promoters that have low signal in *Dnmt1*$^{+/+}$ MEFs show a greater variation in *Dnmt1*$^{-/-}$ MEFs with a tendency for K4me3 gain (Fig. 4.2E upper and lower). This suggests that many promoters that are not normally marked by K4me3 are dependent on an intact DNA methylation system to prevent accumulation of this mark.

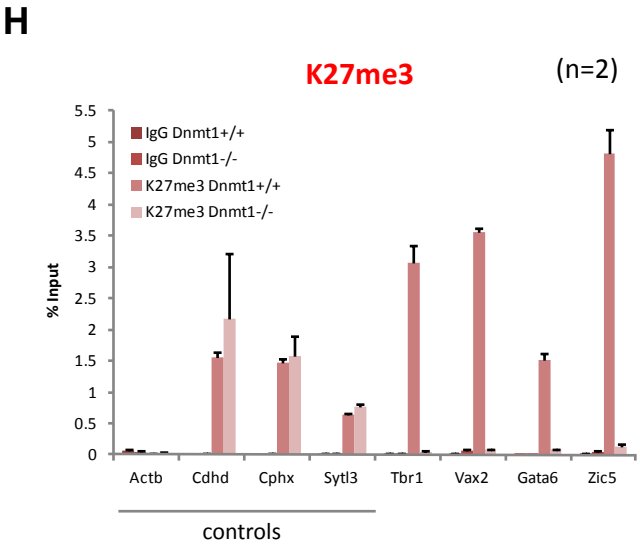
Among the list of K27me3 loss genes (promoters that show decreased K27me3 in *Dnmt1*$^{-/-}$ MEFs below the set threshold) were many genes with known roles in embryonic development. Among these genes were representatives from each of the four *Hox* clusters. The microarray signal at *HoxC* promoters was checked to make sure that it matches the results of N-ChIP-qPCR. A large enrichment of K27me3 is observed at *HoxC* promoters in *Dnmt1*$^{+/+}$ and this enrichment is lost in *Dnmt1*$^{-/-}$ MEFs while increased K4me3 is observed for *Hoxc9-13*, confirming the findings by N-ChIP-qPCR (Fig. 4.3B). Interestingly, similar results were observed for the *HoxA* and *HoxD* cluster promoters with decreased K27me3 observed throughout the clusters and K4me3 increases observed at *Hox9-13* (Fig. 4.3A and C). The *HoxB* cluster showed a different pattern with strongly decreased K27me3 observed at *Hoxb2-8* promoters and no increases observed for K4me3 (Fig. 4.3D). Many promoters of genes not in *Hox* clusters also showed striking loss of K27me3 in *Dnmt1*$^{-/-}$ MEFs, as illustrated by the *Vax2*, *Tbr1* and *Zic5* promoters (Fig. 4.3E to G). To show that many PcG targets are not affected and to validate some of the losses of K27me3 at developmental gene promoters, N-ChIP-qPCR was performed. In agreement with the microarray, K27me3 at *Cdhd*, *Cphx* and *Sytl3* promoters was unchanged in *Dnmt1*$^{-/-}$ MEFs while at *Tbr1*, *Vax2* *Gata6* and *Zic5* striking loss of K27me3 was observed (Fig. 4.3H).

Figure 4.3 - Examples of changes in promoter histone modification status in *Dnmt1*$^{-/-}$ MEFs

A to D. Promoter N-ChIP-chip data displayed in the UCSC browser for H3K27me3 and H3K4me3 in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs for *Hox* clusters. Location of RefSeq genes are shown above in blue. Dashed horizontal line indicates y=0. Plots represent the mean normalised log2 ratio of IP over input for each probe over three replicates (see Methods). E to G. As 'A to D' but for *Vax2*, *Tbr1* and *Zic5* promoters. H. N-ChIP-qPCR for examples of non-changing and changing promoters for H3K27me3 in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars represent standard error of the mean (n=2).




E
Vax2
Vax2os2
Vax2os1
Vax2
F
Tbr1
Tbr1
G
Zic5
Zic5
K27me3 Dnmt1+/+
K27me3 Dnmt1-/-
K4me3 Dnmt1+/+
K4me3 Dnmt1-/-

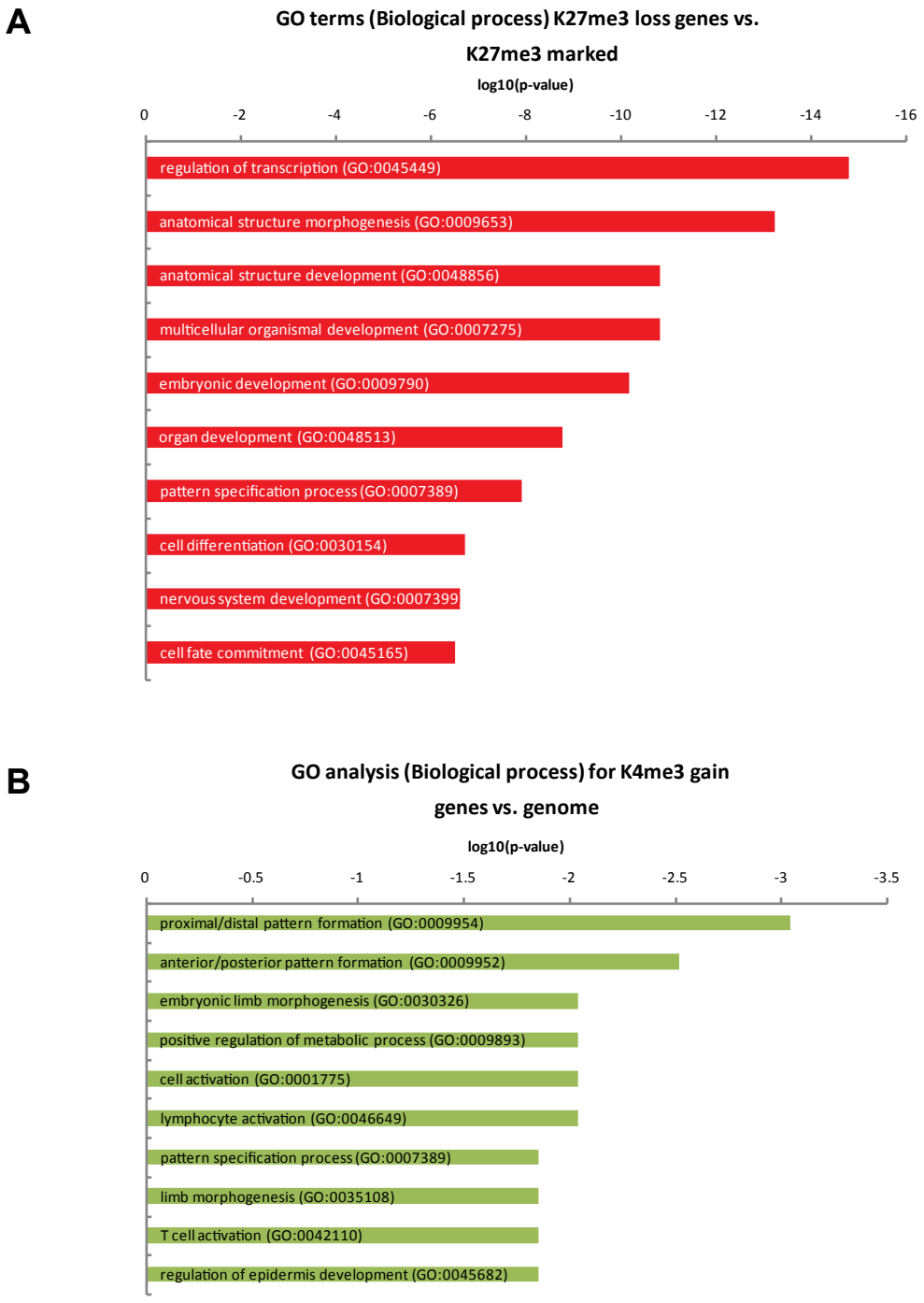

H
K27me3
(n=2)
IgG Dnmt1+/+
IgG Dnmt1-/-
K27me3 Dnmt1+/+
K27me3 Dnmt1-/-
% Input
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Actb
Cdhd
Cphx
Sytl3
Tbr1
Vax2
Gata6
Zic5
controls

To test for an enrichment in functions in embryonic development among genes that show altered histone marks in *Dnmt1*$^{-/-}$ MEFs, Gene Ontology (GO) analysis was performed. No significant enrichments were observed for K27me3 gain genes or K4me3 loss genes (data not shown). K27me3 loss genes were highly enriched for GO terms involved in embryonic development and functions in transcriptional regulation (Fig. 4.4A, top 10 terms are shown). Importantly, as PcG proteins are known to mark many genes with functions in development, this enrichment analysis was performed against a list of genes marked by K27me3 in *Dnmt1*$^{+/+}$ MEFs (see Methods for details). The same terms were found to be significantly enriched when tested against the genome as background (data not shown). K4me3 gain genes showed significant enrichments for GO terms related to certain aspects of embryonic development such as body axis pattern formation (Fig. 4.4B, top 10 terms are shown). Both K27me3 loss and K4me3 gain genes showed enrichment for GO terms that are attached to *Hox* genes, enrichment for the interpro protein domain homeobox (HOX) and nuclear localisation of the encoded protein (data not shown). Overall, GO analysis suggested that both of these gene sets are strongly enriched for transcription factors that have important functions in embryonic development. By way of confirmation, three genes were chosen that encode proteins with key roles in embryonic development, and show K27me3 loss in *Dnmt1*$^{-/-}$ MEFs by ChIP-chip. *Vax2* and *Hand2* are transcription factors with important roles in morphogenesis of the eye and various organs respectively (Alfano *et al.*, 2011; Galli *et al.*, 2010). *Wnt10a* is a secreted morphogen involved in the development of multiple structures through the Wnt-signalling pathway (Wang and Shackleford, 1996). N-ChIP-qPCR confirmed that these promoters lose K27me3 in *Dnmt1*$^{-/-}$ MEFs (Fig. 4.4C). To check if K27me3 loss at these genes correlates with increased expression in hypomethylated cells I performed qRT-PCR in *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ MEFs and in NIH3T3 MEFs treated with 5-aza-dC. All three genes showed increased expression in hypomethylated cells (Fig. 4.4D and E).

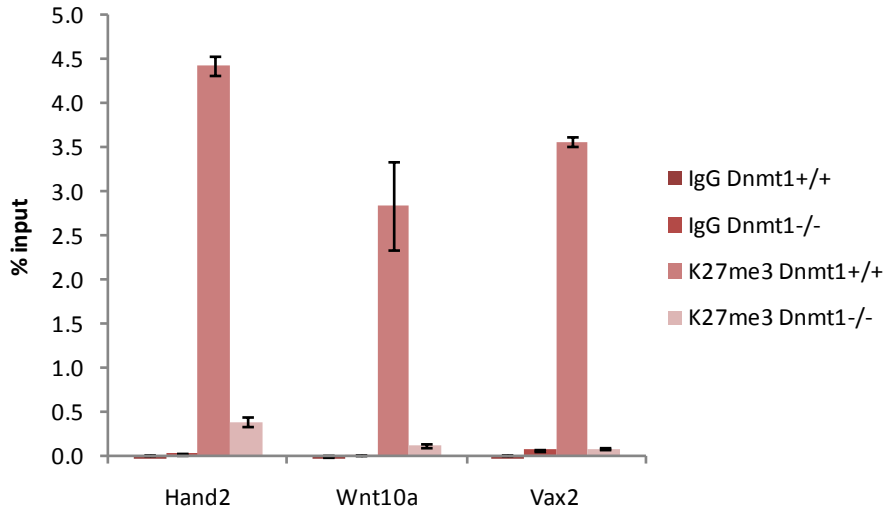
In summary, in *Dnmt1*$^{-/-}$ MEFs a large number of gene promoters have altered histone modifications. The bulk of K27me3 changes involve loss of the modification at genes that are normally highly K27me3 marked, many of which have key functions in embryonic development. The bulk of K4me3 changes involve gain of the modification at genes that are normally not marked by K4me3, many of which have key functions in embryonic development.

Figure 4.4- Many genes with changing histone modification status in *Dnmt1*$^{-/-}$ MEFs have functions in development

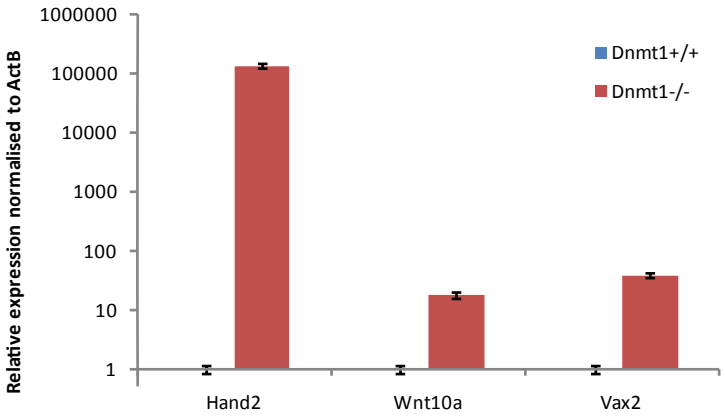
A and B. GO enrichment analysis (see Methods) of genes that lose K27me3 or gain K4me3 respectively in *Dnmt1*$^{-/-}$ MEFs. Plotted are the Bonferroni corrected log10 transformed p-values for enrichment as calculated using Babelomics4. C. N-ChIP-qPCR for K27me3 at the promoters of selected developmental genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars indicate standard error of the mean. D and E. Relative expression of selected developmental genes measured by qRT-PCR in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs and AZA treated relative to control MEFs respectively. Error bars represent standard error of the mean.



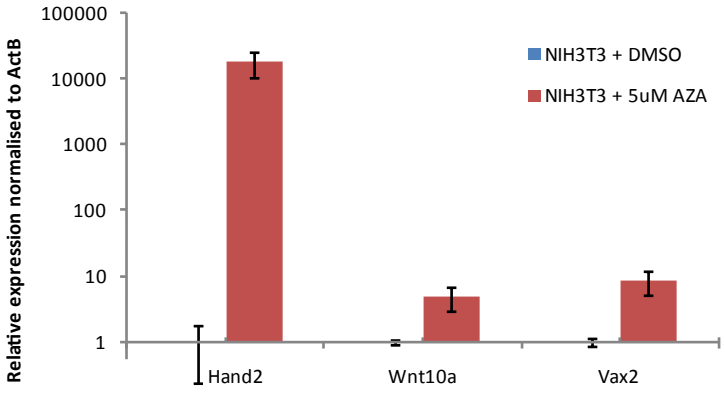
C


5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
% input
Hand2
Wnt10a
Vax2
IgG Dnmt1+/+
IgG Dnmt1-/-
K27me3 Dnmt1+/+
K27me3 Dnmt1-/-

D


1000000
100000
10000
1000
100
10
1
Relative expression normalised to ActB
Hand2
Wnt10a
Vax2
Dnmt1+/+
Dnmt1-/-

E


100000
10000
1000
100
10
1
Relative expression normalised to ActB
Hand2
Wnt10a
Vax2
NIH3T3 + DMSO
NIH3T3 + 5uM AZA

4.2.3. The relationship between changes in promoter histone modifications and gene expression

In order to assess the relationship between changes in K27me3 and K4me3 at gene promoters and gene expression, I integrated my promoter N-ChIP-chip data with data from an expression microarray performed on the same cells (performed by J.Hackett). Firstly, to make sure the expected patterns were present, I plotted the signal for each histone modification for genes at different expression levels. As expected, promoter K27me3 was more abundant at genes with low expression levels and the inverse was true for K4me3 (Fig. 4.5A and B). As K27me3 is thought to be involved in PcG-mediated gene silencing I plotted the relative expression of genes that lose and gain K27me3 in *Dnmt1*$^{-/-}$ MEFs. Generally, genes that lose promoter K27me3 are upregulated in *Dnmt1*$^{-/-}$ MEFs while genes that gain K27me3 showed no change (Fig. 4.5C). K4me3 is thought to be involved in gene activation through transcription initiation (Kouzarides, 2007). Generally, genes that gain K4me3 are also upregulated in *Dnmt1*$^{-/-}$ MEFs while genes that lose K4me3 are downregulated (Fig. 4.5D). When K27me3 loss and K4me3 gain genes are compared, a greater increase in gene expression is observed for K4me3 gain genes than K27me3 loss genes overall (Fig. 4.5E). Genes that both lose K27me3 and gain K4me3 show a further increase in expression suggesting that these two histone marks function at least partially independently (Fig. 4.5E). These results alone suggest that changes in histone marks are associated with changes in gene expression. The small number of genes that show increased K27me3 in *Dnmt1*$^{-/-}$ MEFs are generally not downregulated at the level of gene expression indicating that they are not *de novo* repressed by PcG proteins.

In order to begin to address the mechanism that drives the histone mark changes in *Dnmt1*$^{-/-}$ MEFs I investigated their relationship with gene expression changes in more detail. One explanation for these changes is that DNA methylation, or Dnmt1, loss leads to changes in transcription which then influences these marks. To test this possibility a heatmap was plotted for K27me3, K4me3 and gene expression changes for all genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs (Fig. 4.6A). Hierarchical clustering of genes was performed which separated them into six major groups based on these characteristics (gene groups are summarised in Table 4.1 below). Of note, genes showing the most K27me3 loss and K4me3 gain were clustered into two groups, groups 1 and 2 (Fig. 4.6A, Table 4.1). Group 1 genes are upregulated at the level of gene expression while group 2 genes show no change in expression (Fig. 4.6A and D, Table 4.1). Interestingly, group 2 genes show the greatest loss of K27me3 and the greatest gain of K4me3 of all the six groups despite no change in

Figure 4.5 - Relationship between histone modification changes and gene expression changes

A and B. Boxplots showing signal of K27me3 and K4me3 respectively, as measured by promoter ChIP-chip, in *Dnmt1*+/+ MEFs of genes split into deciles based on their expression in *Dnmt1*+/+ MEFs. Red spots indicate the mean of each group. C and D. Boxplots showing log2 relative expression of genes in *Dnmt1*-/- relative to *Dnmt1*+/+ MEFs. Genes are split into groups depending on their histone modification change in *Dnmt1*-/- MEFs as measured by ChIP-chip (C shows groups based on K27me3 changes, D shows groups based on K4me3 changes, E shows groups based on both K4me3 and K27me3 changes). P-values are from Mann-Whitney U tests between indicated groups.

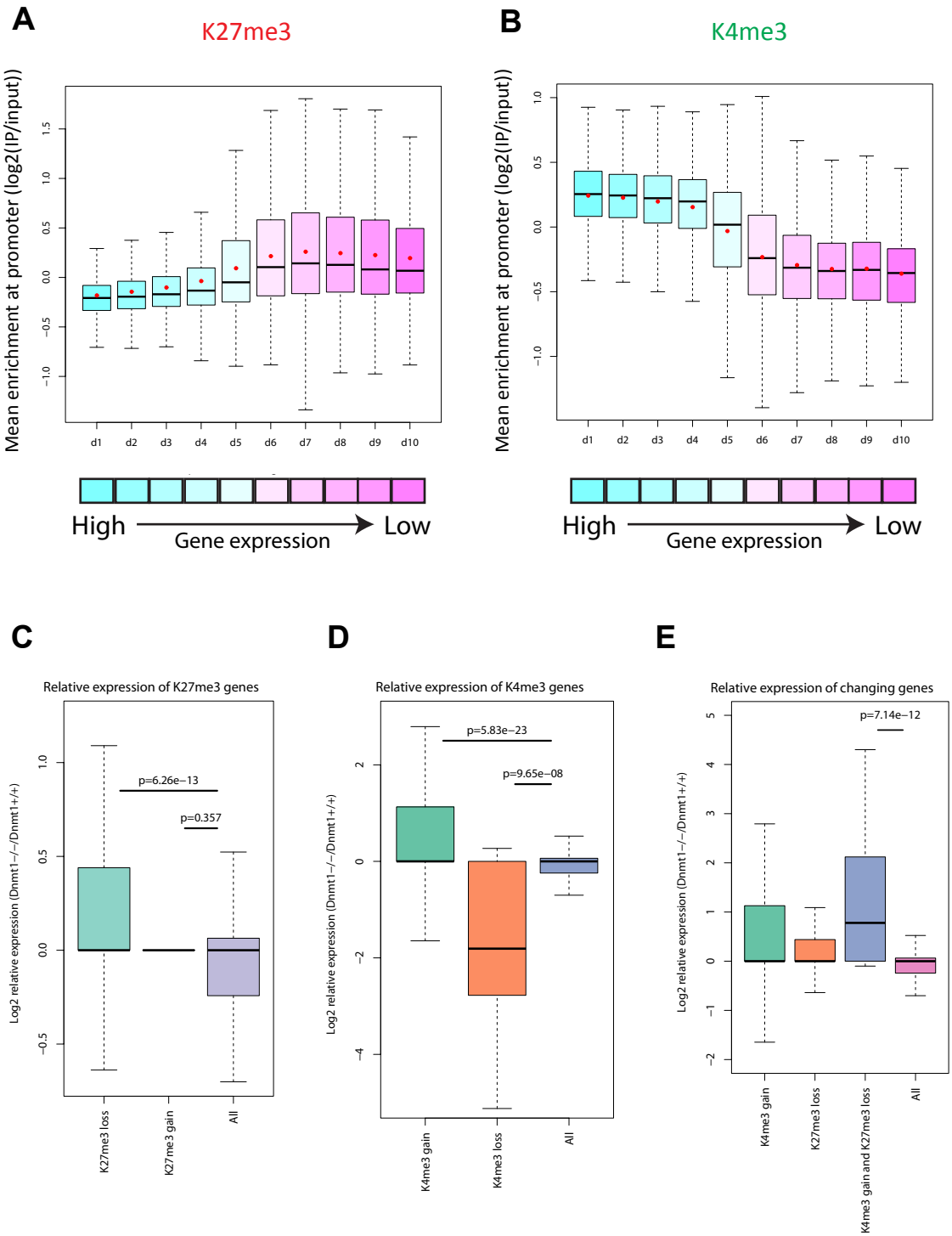
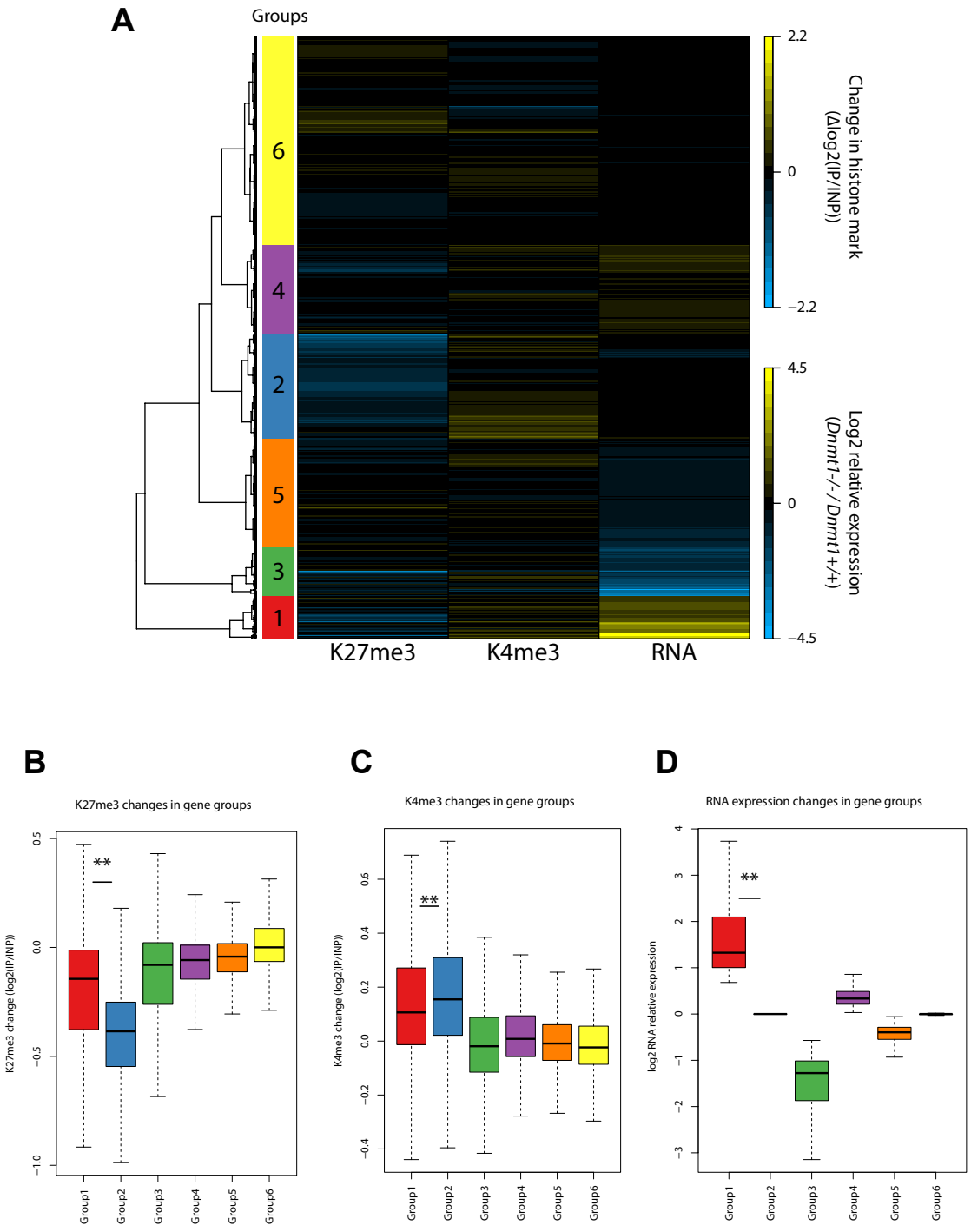
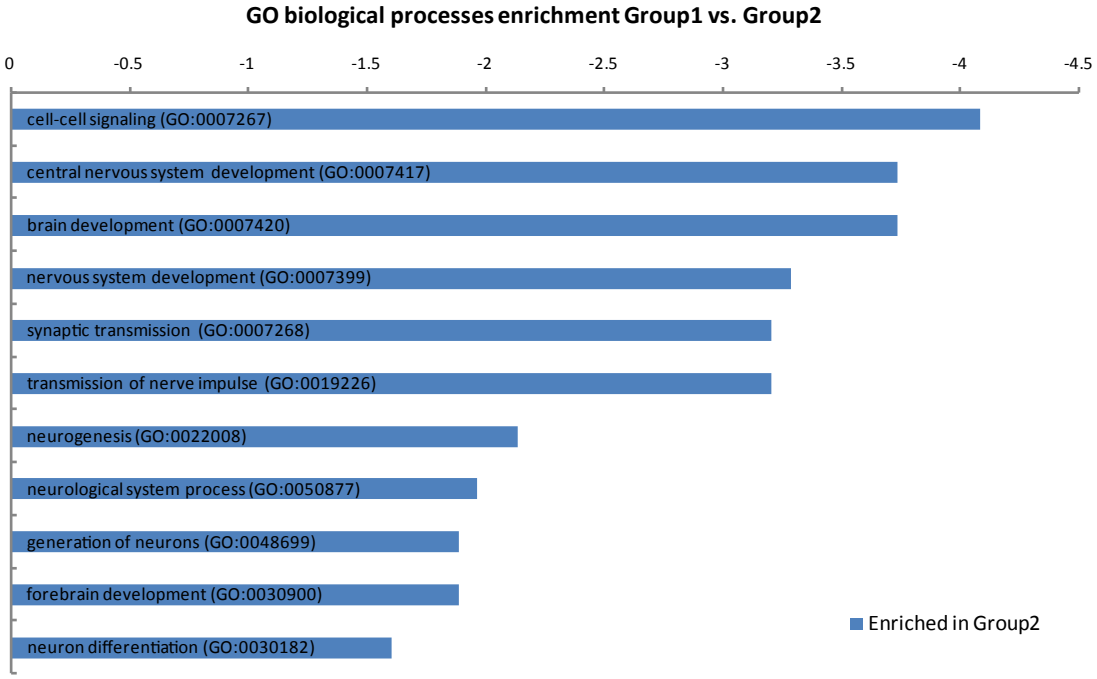


Figure 4.6 - Heatmap and hierarchical clustering of genes shows uncoupling of histone marks and gene expression

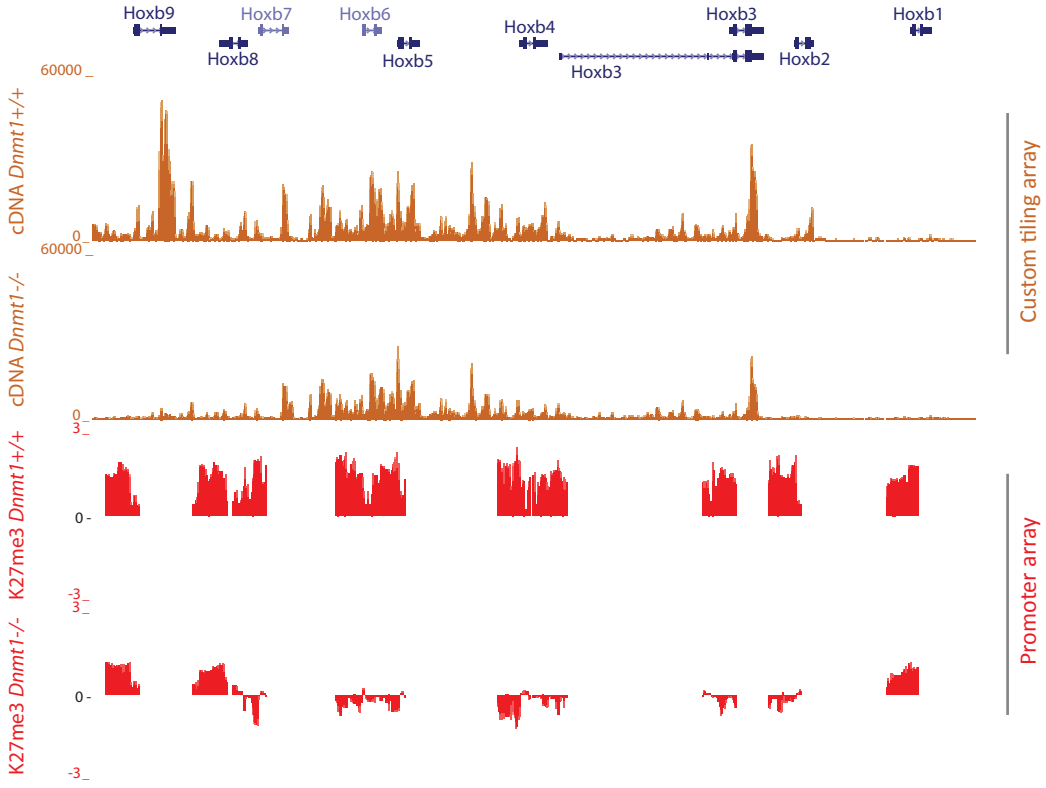
A. Heatmap showing changes in K27me3, K4me3 and RNA expression in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs (see Methods). Dendogram shows results of hierarchical clustering by which genes were classified into 6 groups (see Methods). B, C and D. Boxplots showing K27me3, K4me3 and RNA expression changes respectively of the groups classified in 'A'. P-values are from Mann-Whitney U tests between the indicated groups. E. Selected GO terms that are significantly enriched in group 2 relative to group 1. F. UCSC browser image displaying data for the *HoxB* cluster from K27me3 N-ChIP-chip (promoter array) and cDNA microarray (custom tiling array) experiments in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Position of Refseq genes (mm9) is shown above in blue.



E


GO biological processes enrichment Group1 vs. Group2
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
cell-cell signaling (GO:0007267)
central nervous system development (GO:0007417)
brain development (GO:0007420)
nervous system development (GO:0007399)
synaptic transmission (GO:0007268)
transmission of nerve impulse (GO:0019226)
neurogenesis (GO:0022008)
neurological system process (GO:0050877)
generation of neurons (GO:0048699)
forebrain development (GO:0030900)
neuron differentiation (GO:0030182)
Enriched in Group2

F


Hoxb9
Hoxb7
Hoxb6
Hoxb4
Hoxb3
Hoxb1
Hoxb8
Hoxb5
Hoxb3
Hoxb2
60000
cDNA Dnmt1+/+
0
60000
cDNA Dnmt1-/-
Custom tiling array
3
0
K27me3 Dnmt1+/+
-3
3
0
K27me3 Dnmt1-/-
-3
Promoter array

expression (Fig. 4.6A,B,C and D, Table 4.1). This suggests that K27me3 loss and K4me3 gain can be uncoupled from transcriptional changes. This is important for the following reasons; 1) it suggests that changes in gene transcription do not cause all of the observed changes in histone marks, and, 2) it suggests that changes in histone marks are not sufficient to cause changes in gene expression. In order to investigate the difference between group 1 (upregulated) and group 2 (not upregulated) genes I looked for enrichments in GO terms between the groups (Fig. 4.6E). No terms were found to be enriched in group 1 over group 2, including a search for transcription factor motifs (data not shown). Many terms were enriched in group 2 over group 1, most of which were related to neural development (Fig. 4.6E). This suggests that many genes that lose K27me3 and gain K4me3 but are not transcriptionally upregulated have functions in neural development. An example of such genes are the *Hoxb2-7* genes which lose K27me3 in *Dnmt1*$^{-/-}$ MEFs but are not transcriptionally upregulated and, in fact some are downregulated (Fig. 4.6F). These results suggest that transcriptional changes are not upstream of changes in promoter K27me3 and K4me3 in this system.

Group (defined by hierarchical clustering)	General K27me3 change in *Dnmt1*$^{-/-}$	General K4me3 change in *Dnmt1*$^{-/-}$	General RNA change in *Dnmt1*$^{-/-}$	Number of genes in group	Example genes
1	- - -	+++	+++	1021	*Hoxc9, Hoxc13, Vax2*
2	- -	+++	none	2472	*Gata1, Foxn1, Wnt8a*
3	-	none	---	1135	*Lefty1, Hoxb9, Wnt4*
4	none	none	+	2094	*Magea9, Tex24, Klf1*
5	none	none	_	2550	*Twist1, Ets2, Med12*
6	none	none	none	4891	*Taf1b, Suz12, Bmi1*

Table 4.1 - Summary of gene groups defined by hierarchical clustering in Fig. 4.6. + and - indicate increase and decrease respectively. Multiple + or - (e.g. +++) indicates a greater difference.

4.2.4. Relationship between promoter K27me3 and K4me3 changes

K4me3 has been suggested to inhibit the binding and activity of the PRC2 complex *in vitro* (Schmitges *et al.*, 2011). I reasoned that one potential mechanism for K27me3 loss in *Dnmt1*$^{-/-}$ MEFs is increased K4me3. To test this possibility I examined the relationship between K27me3 and K4me3 changes at gene promoters. Although significantly correlated (by Pearson's correlation test), plotting K27me3 loss and K4me3 gain produced an R value

of 0.283 meaning that only 8% of the variation in K27me3 can be explained by variation in K4me3 at gene promoters (Fig. 4.7A, R^2=0.08). As this plot contains all analysed gene promoters, which may have diluted a larger dependence, I performed the same analysis using only genes that are marked by K27me3 in *Dnmt1*$^{+/+}$ MEFs (Fig. 4.7B). Again although the two variables are significantly correlated, only 2.7% of the variation in K27me3 could be explained by variation in K4me3 (Fig. 4.7B, R^2=0.027). These results suggest that promoter K27me3 and K4me3 changes in *Dnmt1*$^{-/-}$ MEFs are at least partially independent. As an example of independence, the *HoxB* locus is shown (Fig. 4.7C). *Hoxb2-7* show a striking decrease in K27me3 in *Dnmt1*$^{-/-}$ MEFs but no increases in K4me3 are observed. It is of course a possibility that increases in K4me3 outside of promoter regions assayed here are responsible for K27me3 loss or that more subtle changes are involved that were missed by averaging across the whole promoter region.

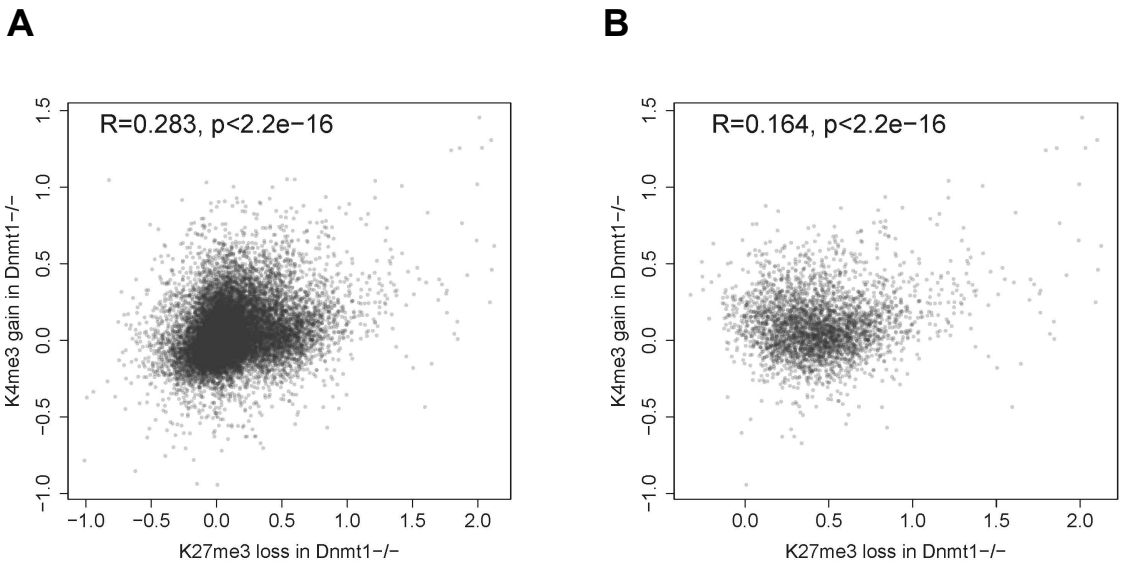
4.2.5. Relationship between histone mark changes and DNA methylation at promoters

In order to investigate what the mechanism behind the changes in promoter histone marks is in *Dnmt1*$^{-/-}$ MEFs I examined promoter DNA methylation levels. First, as promoter CpG content has been linked to both DNA methylation and histone modifications, I assessed the promoter structure of different groups of genes (Weber *et al.,* 2007; Mikkelsen *et al.,* 2007). K27me3 loss genes are enriched for HCPs (High CpG content Promoters), with almost 80% having a HCP, and are depleted for both ICPs (Intermediate CpG content Promoters) and LCPs (Low CpG content Promoters)(Fig. 4.8A). Importantly, this comparison was made with K27me3 marked genes in *Dnmt1*$^{+/+}$ MEFs to control for any bias of PcG marked genes. K27me3 gain genes are enriched for ICPs and LCPs and depleted for HCPs when compared to all genes (Fig. 4.8A). Both K4me3 gain and K4me3 loss genes are enriched for ICPs and depleted for HCPs (Fig. 4.8B). These results are interesting as a very high proportion of HCPs are thought to be unmethylated whereas ICPs are thought to be methylated more frequently and show a greater variation in methylation between tissues (Weber *et al.,* 2007).

In order to map promoter methylation in *Dnmt1*$^{+/+}$ MEFs, we established a collaboration with the lab of J.Greally (Albert Einstein College of Medicine, NY) who performed HELP-tag-seq (HpaII tiny fragment Enrichment by Ligation-mediated PCR followed by Tag Sequencing). HELP-tag-seq involves digestion of genomic DNA with the isoschizomers HpaII (methylation sensitive) and MspI (not methylation sensitive) followed

Figure 4.7 - Relationship between K27me3 and K4me3 changes at promoters

A. Scatter plot showing relationship of changes in histone modifications (at the promoter level) at all assayed promoters in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs as measured by N-ChIP-chip on promoter arrays. Scale on each axis is the difference in log2(IP/INP)(see Methods) for each histone modification between *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs. R-value and p-value for correlation were calculated using a Pearson's product moment test. B. As 'A' but using only genes that are marked by K27me3 in *Dnmt1*$^{+/+}$ MEFs (log2(IP/INP) of > 0.5). C. UCSC browser image showing promoter N-ChIP-chip data at the *HoxB* cluster for *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Copied from figure 4.3D.



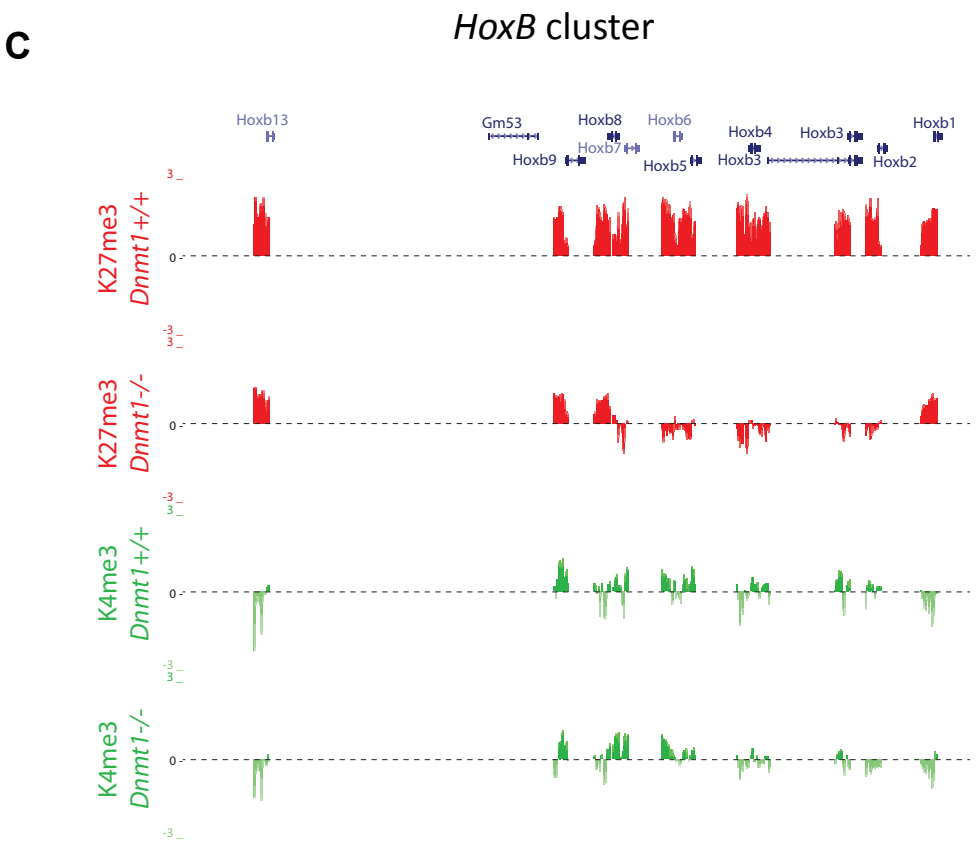
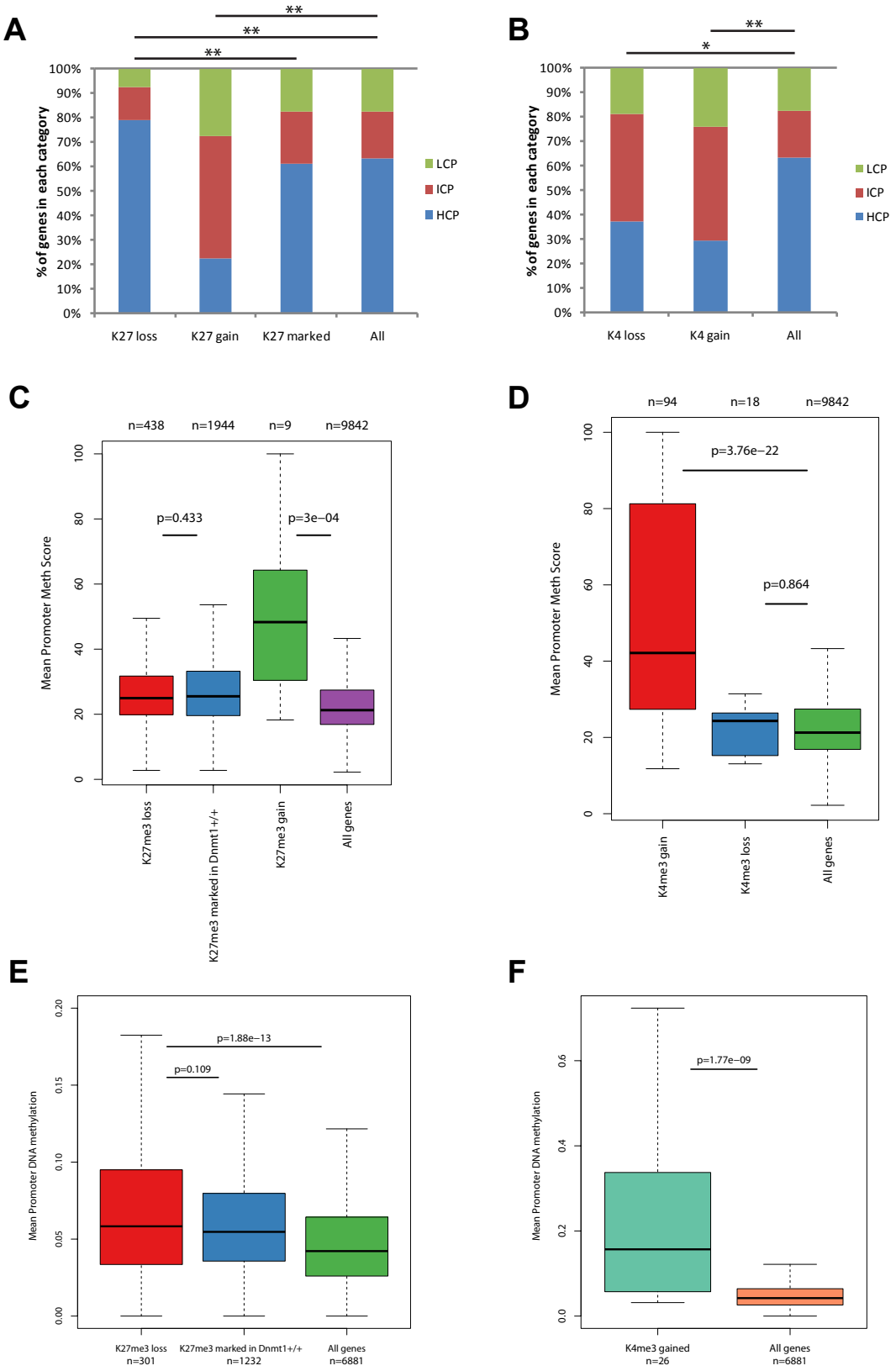


Figure 4.8 - Promoter structure and DNA methylation

A and B. Promoter CpG content classified as HCP (High CpG density promoter), ICP (Intermediate) or LCP (Low) (according to Mikkelsen *et al.,* 2007) for K27me3 and K4me3 gene categories respectively. ** and * represent p<0.001 and p<0.01 respectively by chi-squared test between indicated categories. C and D. Promoter DNA methylation level as measured by HELP-seq for K27me3 and K4me3 gene categories respectively. P-values calculated by Mann-Whitney U test between indicated groups. N-values indicate the number of genes in each category. E and F. As 'C and D' but using published promoter DNA methylation data for MEFs (Meissner *et al.,* 2007).

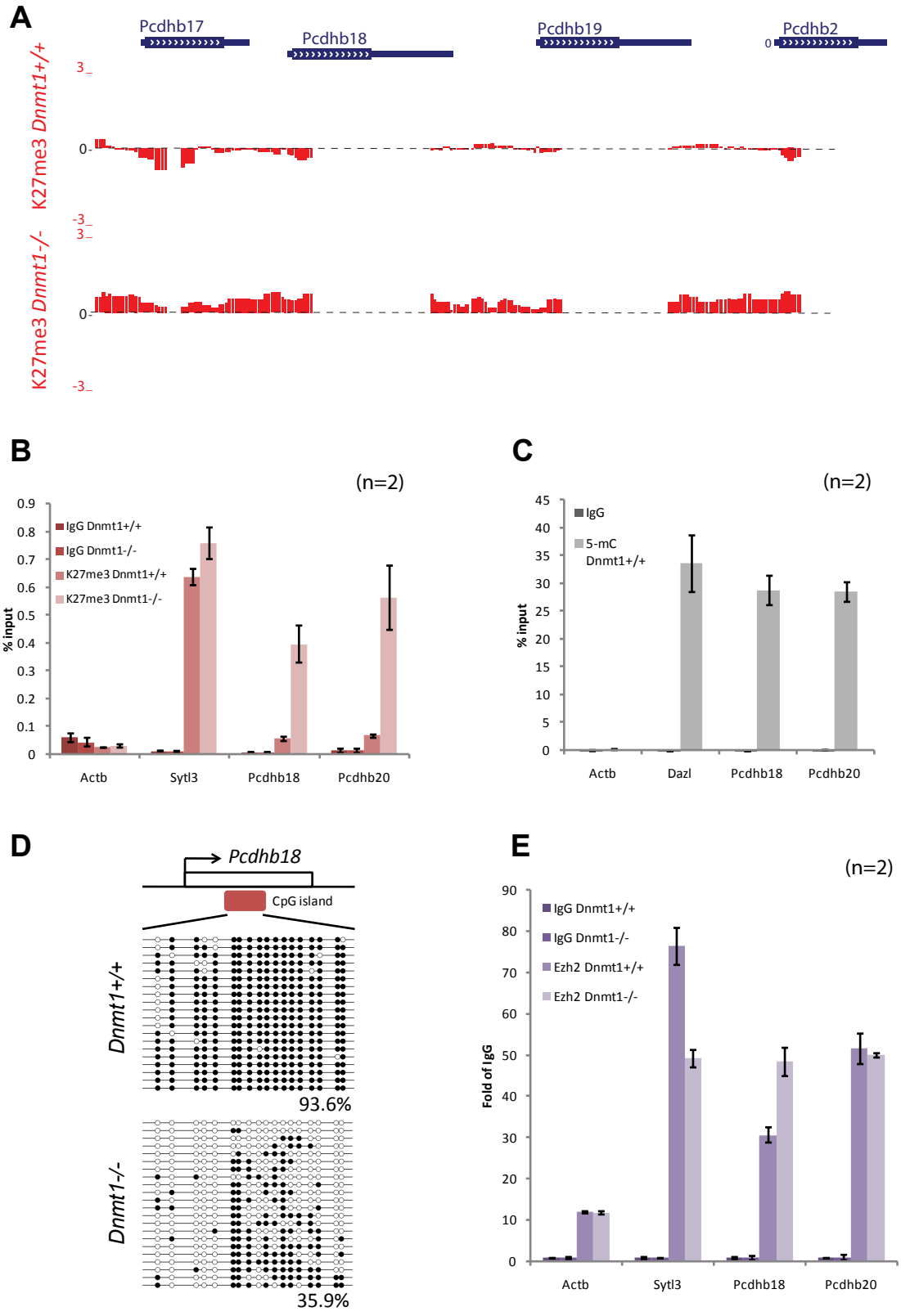


by sequencing of the ends by Illumina sequencing (see Methods for full description)(Khulan *et al.*, 2006; Oda *et al.*, 2009; Suzuki *et al.*, 2010). Depending on sequence coverage, this method allows detection of methylation at potentially >1 million CCGG sites in the mouse genome. See chapter 2 for data processing and validation of HELP-tag-seq data. Chapter 2 suggests that this procedure, which is still in development, is associated with a considerable amount of noise, at least at the sequence depth achieved here. For this reason, the HELP-tag-seq data here should be interpreted with caution. DNA methylation at regions of interest, *ox* gene clusters, is assessed in further detail in chapter 5. Nevertheless, the DNA methylation status of 13,095 promoters was estimated using this HELP-tag-seq at the sequence depth achieved. Using this method, K27me3 loss genes do not have a higher promoter DNA methylation level in *Dnmt1*$^{+/+}$ MEFs than genes marked by K27me3 (Fig. 4.8C). K27me3 gain genes have more DNA methylation at their promoters than all genes, although the number of these genes where DNA methylation data is available is low (Fig. 4.8C). K4me3 gain genes have more promoter DNA methylation than all genes while K4me3 loss genes are not different from all genes (Fig. 4.8D). Importantly, these trends were repeated using published promoter DNA methylation data (reduced representation bisulfite sequencing) for wildtype MEFs (Fig. 4.8E and F)(Meissner *et al.*, 2008). K27me3 gain and K4me3 loss genes are not represented in this data so could not be included in this analysis. These observations together suggest that K27me3 loss occurs mainly at HCPs that have low DNA methylation, and K27me3 gain and K4me3 gain occurs mainly at ICPs that have higher than average DNA methylation.

DNA methylation has been suggested to prevent K27me3 targeting to promoter proximal regions of some genes in neural stem cells (Wu *et al.*, 2010). A similar process may be involved in K27me3 increases at K27me3 gain genes in *Dnmt1*$^{-/-}$ MEFs. As K27me3 gain genes have more promoter DNA methylation than genes on average, loss of this methylation in *Dnmt1*$^{-/-}$ MEFs could lead to increased K27me3. To study this relationship I chose to investigate the *Pcdhb* cluster which shows increased K27me3 in *Dnmt1*$^{-/-}$ MEFs by promoter N-ChIP-chip (Fig. 4.9A). I confirmed increases in K27me3 at two genes of this cluster, *Pcdhb18* and *Pcdhb20*, by N-ChIP-qPCR while a control gene, *Sytl3*, showed no major change (Fig. 4.9B). Both of these genes contain an intragenic CpG island. To assess the DNA methylation status of the intragenic CpG islands I performed MeDIP-qPCR in *Dnmt1*$^{+/+}$ MEFs (Fig. 4.9C). Both of these loci showed a large enrichment for methylated DNA relative to CpG-rich negative (*Actb*) and positive (*Dazl*) controls suggesting that they are highly methylated. This was confirmed for the intragenic CpG island of *Pcdhb18* by bisulfite sequencing (Fig. 4.9D). This region in *Dnmt1*$^{-/-}$ MEFs showed markedly reduced

Figure 4.9 - Candidate K27me3 gain genes have high levels of DNA methylation

A. UCSC browser image showing K27me3 promoter ChIP-chip data at the *Pcdhb* gene cluster in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. B. ChIP-qPCR for K27me3 at control and *Pcdhb* genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars indicate standard error of the mean. C. MeDIP-qPCR for control and *Pcdhb* genes in *Dnmt1*$^{+/+}$ MEFs. D. Bisulfite sequencing for the *Pcdhb18* intragenic CpG island in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. E. ChIP-qPCR for Ezh2 in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs at control and *Pcdhb* genes. Error bars indicate standard error of the mean.



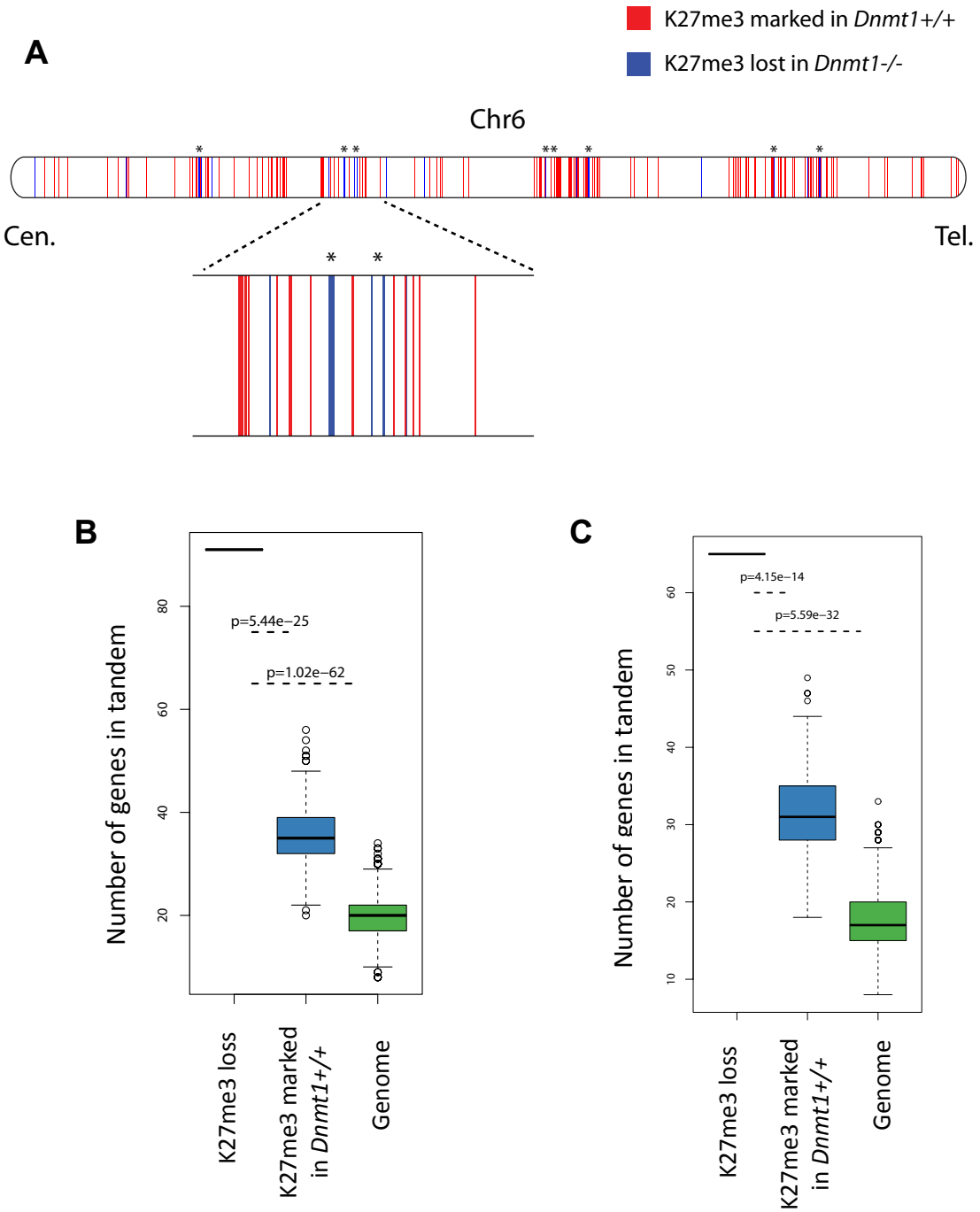
DNA methylation (Fig. 4.9D). In order to assess if increased K27me3 is caused by increased binding of the PRC2 complex to this CpG island, X-ChIP (cross-linked ChIP) for Ezh2 was performed (Fig. 4.9E). Increased Ezh2 binding was observed at the intragenic CpG island of *Pcdhb18* in *Dnmt1*$^{-/-}$ MEFs but not at the intragenic CpG island of *Pcdhb20* or a control gene (Fig. 4.9E). Together, these data suggest that candidate K27me3 gain genes contain highly methylated CpG-rich regions and, at one tested region, increased binding of the PRC2 complex occurs when it is hypomethylated in *Dnmt1*$^{-/-}$ MEFs.

4.2.6. Genomic organisation of K27me3 loss genes

I noticed that many K27me3 loss genes are next to each other in the genome. This included genes at all four *Hox* clusters and other genes that are located in genomic clusters. To test if this observation represents a general characteristic of K27me3 genes I investigated the genomic location of K27me3 genes. By plotting chromosome ideograms it was apparent that many genes that lose K27me3 in *Dnmt1*$^{-/-}$ MEFs exist in clusters in the genome or are neighboured by genes that also lose K27me3 (Fig. 4.10A). To calculate if this represents a statistically significant characteristic I counted the number of pairs of K27me3 genes that occur in tandem in the genome (Fig. 4.10B)(see Methods for details). In order compare this number to what would be expected by chance, I sampled genes randomly from the genome and counted the number of pairs over 1000 permutations (Fig. 4.10B). I also did this for genes that are K27me3 marked to control for an effect of being a PcG target. The number of gene pairs in K27me3 loss genes is significantly higher than what would be expected by chance (Fig. 4.10B). To exclude that the *Hox* clusters were causing this effect, the analysis was repeated using a genome where the genes from all four *Hox* clusters had been removed (Fig. 4.10C). These results suggest that K27me3 loss in *Dnmt1*$^{-/-}$ MEFs may frequently occur over large genomic domains encompassing multiple genes, as is observed at the *Hox* clusters.

Figure 4.10 - K27me3 loss genes are found in tandem in the genome more often than would be expected by chance

A. Chromosome ideogram of chr6 showing positions of gene promoters marked by K27me3 in *Dnmt1*$^{+/+}$ (red) and gene promoters that lose K27me3 in *Dnmt1*$^{-/-}$ (blue) as measured by N-ChIP-chip (promoter array). * indicates positions of blue gene promoters that are in tandem (see Methods). Note that gene promoters in close proximity appear as a thicker blue line. Centomeric (Cen.) and telomeric (Tel.) ends of the chromosome are marked. A portion of chr6 is magnified below. B. Boxplot showing number of genes found in tandem in K27me3 loss genes (one value) and randomly selected genes from K27me3 marked genes and all genes (1000 samples - see methods). C. As 'B' but removing *Hox* genes from all lists.



4.3. Discussion

The results presented in this chapter suggest that an intact DNA methylation system is required for proper localisation of the histone marks K27me3 and K4me3. The implications of these findings for our understanding of the roles of DNA methylation in genome regulation are discussed in this section under the following headings.

4.3.1. Is an intact DNA methylation pattern required for PRC2 targeting?

How the PRC2 complex is targeted to genomic regions is unknown and represents an outstanding problem in the field (Schuettengruber and Cavalli, 2009). The results presented in this chapter implicate an intact DNA methylation system in this process and could therefore inform us about PRC2 targeting in mammals. The most surprising observation in this chapter is that decreased K27me3, a signature of PRC2 binding, is observed at many gene promoters in severely DNA hypomethylated *Dnmt1*$^{-/-}$ MEFs, but globally, K27me3 is not decreased. This suggests that the K27me3 mark is not globally decreased in *Dnmt1*$^{-/-}$ MEFs but instead is redistributed to different regions of chromatin. The genes that lose K27me3 in *Dnmt1*$^{-/-}$ MEFs are among the most highly enriched for K27me3 in *Dnmt1*$^{+/+}$ MEFs and the vast majority are associated with a CpG-rich promoter that contains low levels of DNA methylation. This is consistent with a redistribution of the K27me3 mark away from its normal targets. I think that this model best explains the observations of this chapter and it will be discussed briefly here and in more detail in chapter 6.

Recently, studies of PRC2-targeting in mammals have implicated CpG islands in the recruitment of the PRC2 complex (Fig. 1.7i)(Tanay *et al.*, 2007; Ku *et al.*, 2008; Mendenhall *et al.*, 2010). DNA methylation has also been suggested to have a negative effect on the binding and/or activity of the PRC2 complex on chromatin (Bartke *et al.*, 2010; Wu *et al.*, 2010; Lynch *et al.*, 2011). It is possible that DNA methylation is required to differentiate CpG island chromatin from surrounding regions and also from DNA methylated CpG islands. In severely DNA hypomethylated cells the PRC2 complex may therefore be titrated from its normal binding sites by *de novo* binding to previously DNA methylated regions (discussed further in chapter 6). In a recent publication it was suggested that K27me3 is found in ectopic peaks in mouse ES cells lacking Dnmt3a and Dnmt3b (Lynch *et al.*, 2011). Many of these ectopic peaks are found in CpG-rich regions that are highly methylated in

wildtype ES cells, supporting this idea. However, in this study, only modest reduction in K27me3 from *Hox* gene clusters was observed, in contrast to the large reduction in this mark observed at *Hox* gene clusters in *Dnmt1*$^{-/-}$ MEFs (Lynch *et al.*, 2011). Possible reasons for the discrepancy between the two different hypomethylated cell types will be discussed in chapter 6. These observations together suggest that the DNA methylation pattern has a major influence on the genomic binding of the PRC2 complex, and in this way may influence the expression of many developmentally important genes. For example, tissue-specific differences in DNA methylation may control the expression of genes through modulation of PRC2 binding to regions of chromatin. Tissue-specific DNA methylation of CpG islands may be important here.

It is clear from my data that certain gene promoters show larger reductions in K27me3 than others. Among these genes are the developmentally important *Hox* genes. How might these differences arise? Are there additional factors that influence K27me3 loss at these genes? The processes that have been implicated in PRC2 targeting and relief from PRC2 binding were shown in figure 1.7. Locus-specific loss of PRC2 binding could therefore arise by altering any of these putative mechanisms. DNA binding proteins have been implicated in PRC2 targeting through binding to polycomb response elements (PREs)(Fig. 1.7i), although it is unclear how much this process is operative in mammals as no mammalian DNA binding protein has yet been identified that clearly performs a similar function. It is possible that proteins required for PRC2 targeting to specific regions are mis-regulated in *Dnmt1*$^{-/-}$ MEFs which causes the observed mis-targeting. The same is true for ncRNA which may be required for targeting of the PRC2 complex to certain regions. Given that not much is known about either of these mechanisms, testing these possibilities would be difficult. K27me3 patterns are also shaped by K27me3 demethylases (Fig. 1.7ii). It is possible that these proteins are differentially targeted in *Dnmt1*$^{-/-}$ MEFs to the genes that lose K27me3. In this respect not much is known about the targeting of these proteins. UTX, a K27me3 demethylase, has a reported interaction with MLL2 and so may be targeted by TrxG complexes (Issaeva *et al.*, 2007).

K27me3 and PRC2 binding is thought to be inhibited by transcription across certain PRC2 binding sites (Fig. 1.7iii)(for example, Schmitt *et al.*, 2005). Loss of PRC2 binding in *Dnmt1*$^{-/-}$ MEFs could conceivably be influenced by intergenic transcription, as DNA methylation is well associated with transcriptional repression and has previously been implicated in reducing transcriptional noise and in regulating alternative promoter usage (Bird, 1995; Skene *et al.*, 2010; Maunakea *et al.*, 2010). Interestingly, abundant intergenic transcription has been reported within *Hox* gene clusters, often correlating with loss of

PRC2-mediated repression of nearby *Hox* genes (Rinn *et al.*, 2007; Sasaki *et al.*, 2007; Sessa *et al.*, 2007). *Hox* clusters contain many orphan CpG islands, some of which have been reported to be DNA methylated in a tissue-specific manner. One hypothesis could be that these orphan CpG islands represent the promoters of ncRNA with functions in activation of local *Hox* genes by PRC2 removal. This idea could be tested by initially examining the DNA methylation status of *Hox* orphan CpG islands in MEFs and the expression of intergenic RNA in *Dnmt1*$^{-/-}$ MEFs, followed by more functional tests.

PRC2 binding and activity has been suggested to be influenced by the presence of other histone modifications (Fig. 1.7iv)(Schmitges *et al.*, 2011; Lau *et al.*, 2011). DNA methylation may be required for the correct patterning of these histone modifications and therefore contribute to PRC2 targeting in this way. One such modification is K4me3 which is reportedly inhibitory to PRC2 binding/catalytic activity *in vitro* (Schmitges *et al.*, 2011). One possibility is that K4me3 accumulates at certain loci in *Dnmt1*$^{-/-}$ MEFs, possibly through increased binding of the methylation-sensitive CXXC-containing Cfp1 or MLL proteins. However, correlating K27me3 and K4me3 changes suggested that this is not the case for the majority of promoter K27me3 changes. This analysis uses only promoter data for K4me3 so it remains possible that increases in K4me3 outside of the regions assayed contribute to K27me3 loss. More detailed analysis of K4me3 changes in *Dnmt1*$^{-/-}$ MEFs would need to be performed to address this possibility fully.

I observed that K27me3 loss genes occur next to each other in the genome much more often than would be expected by chance. This observation is intriguing given the links between PRC2 bound regions and higher order chromatin organisation. PRC2-bound genes are often found within the same nuclear space and reportedly interact with nearby PRC2 marked genes (for example, Bantignies *et al.*, 2011). CTCF is a DNA binding protein implicated in mediating chromatin contacts that is affected by DNA methylation (Philips and Corces, 2009). CTCF has been suggested to interact with PRC2 component Suz12 and has been implicated in targeting PRC2 and also in the regulation of *Hox* genes (Li *et al.*, 2008; Soshnikova *et al.*, 2010; Kim *et al.*, 2011). One model for K27me3 loss could involve differential binding of CTCF resulting in altered chromatin conformation. This could be tested initially by examining the binding of CTCF by ChIP.

Also, mapping of K27me3 in cells treated with 5-aza-dC would help to strengthen this observation by providing a second hypomethylated system. It should be added that at this stage it cannot be excluded that these effects are not mediated by reduction in DNA methylation levels and are a result of loss of Dnmt1 protein itself, as it has suggested non-catalytic functions. In this regard, Dnmt1 has been reported to form a complex with

polycomb protein Bmi1 and Dmap1 (Dnmt1 associated protein 1) and is suggested to be required for proper location of Bmi1 to PRC2 bodies (Hernandez-Munoz *et al.*, 2005; Negishi *et al.*, 2007). Depletion of Dmap1 is also suggested to induce the expression of some PRC2 target genes including certain *Hox* genes (Negishi *et al.*, 2007). Bmi1 is a component of a subset of PRC1 complexes and has not yet been linked to the function of PRC2 and K27me3. Dnmt1 has also been linked, somewhat controversially, to the PRC2 complex directly and an interaction with Ezh2 has been proposed (Vire *et al.*, 2006). The role of these interactions in PRC2 function are yet to be addressed. It would be interesting to try to rescue the K27me3 defects in *Dnmt1*$^{-/-}$ MEFs by expressing wildtype and catalytically inactive Dnmt1. In my opinion, given the published links between 5mCpG and inhibition of PRC2 binding, it seems likely that the DNA methylation mark itself is involved here. Although further investigation of the effect of hypomethylation on PRC2 targeting is clearly required, the results presented in this chapter implicate an intact DNA methylation system in maintaining repression of a large number of developmental genes by somehow facilitating PRC2 targeting to these regions. This represents a putative 'non-canonical' role for DNA methylation in gene regulation as it involves genes that have low levels of DNA methylation at their promoters. It will be interesting to investigate this further, particularly with respect to the molecular mechanism involved.

4.3.2. DNA methylation prevents accumulation of K4me3 at developmental genes?

K4me3 and K4me2 are well associated with regions of high CG-density and low DNA methylation suggesting a functional link between the two epigenetic marks (Weber *et al.*, 2007; Lister *et al.*, 2009; Hawkins *et al.*, 2010; Thomson *et al.*, 2010). Indeed, mapping of DNA methylation and K4me2/3 has shown a strong negative correlation in human cell types (Hawkins *et al.*, 2010). Certain proteins involved in K4me3 deposition contain non-methyl CpG binding CXXC domains and therefore may be involved in this link. One such protein, Cfp1, binds to unmethylated CpG islands in ES cells, interacts with the Set1 H3K4 methylase complex and may be involved in recruitment to these regions (Thomson *et al.*, 2010). The H3K4 methylases MLL1 and MLL2 also contain CXXC domains. The K4me3 mark is thought to be generally associated with transcriptional initiation and is recognised by many proteins involved in establishing a transcriptionally active chromatin state (Kouzarides, 2007). In this way, DNA methylation of promoter regions may act to maintain a repressed state by providing mitotically heritable prevention of transcription initiation signals. To my knowledge, no study has addressed this possibility by investigating the effect

of DNA hypomethylation on K4me3 localisation on a large scale in non-cancer cells. A study of candidate loci in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs revealed that some promoters have increased K4me3 in *Dnmt1*$^{-/-}$ MEFs, warranting a more in depth analysis (Lande-Diner *et al.*, 2007).

In this chapter many gene promoters were found to show increased K4me3 in hypomethylated *Dnmt1*$^{-/-}$ MEFs. In particular, increased K4me3 was observed at promoters that are depleted for this mark in *Dnmt1*$^{+/+}$ MEFs, many of which have high levels of DNA methylation. These genes are enriched for functions in certain aspects of embryonic development. These results support a role for promoter DNA methylation in preventing K4me3 deposition at repressed genes. However, in this thesis I have not focussed on K4me3 changes as I decided to investigate the more novel link with K27me3 in further detail. It would be interesting to analyse my data in more detail in respect to K4me3 changes. One analysis that may be interesting would be to take promoters with high and low DNA methylation and compare data for K4me3 change when hypomethylated in *Dnmt1*$^{-/-}$ MEFs. It may also be interesting to look at non K27me3 targets alone, as K4me3 could conceivably be influenced by changes in K27me3. It would be important to investigate DNA methylation loss in *Dnmt1*$^{-/-}$ MEFs in more detail and correlate this with K4me3 changes. The problem that could be addressed by studying these questions is whether K4me3 gain is universal at normally DNA methylated regions when DNA methylation is lost, or if there is additional specificity involved. We may expect, based on the suggested involvement of CXXC proteins, that the underlying CpG content of a DNA methylated region may influence this effect. Multiple enzymes are thought to be responsible for K4me3 in mammals including five MLL proteins, two Set1 proteins and Ash1 (Kouzarides, 2007). It is possible that recruitment of a subset of these proteins is affected by DNA methylation, and hence more specific K4me3 changes would be observed. MLL proteins have reported specificity, including Mll1 which is thought to be responsible for K4me3 at less than 5% of promoters in MEFs, including *Hox* genes (Wang *et al.*, 2009c). As *Hox* genes are among the genes showing the greatest increase of K4me3 in *Dnmt1*$^{-/-}$ MEFs, and Mll1 contains a CXXC domain, it may be interesting to investigate this protein further. Another possibility is that all DNA methylated regions with sufficient CpG content will show increased K4me3 when hypomethylated. This would suggest a universal effect of DNA methylation on K4me3 prevention, implicating a general non-methyl CpG binding protein, such as Cfp1, in the process. Further analysis outside of gene promoters would be interesting, such as at orphan CpG islands, as these regions are more often DNA methylated than promoter proximal CpG islands (Illingworth *et al.*, 2008; Illingworth *et al.*, 2010). The use of another experimentally

hypomethylated system, such as 5-aza-dC treatment, would compliment this approach. In summary, data presented in this chapter support a role for promoter DNA methylation in K4me3 prevention but more detailed analysis is required to make any further conclusions.

4.3.3. Uncoupling of histone modifications and gene expression

It is open for debate whether histone modifications represent a cause or consequence of a transcription state. By analysing gene expression and histone modification changes in detail, in this chapter I make an interesting observation that these two features can be uncoupled. There are a large number of genes that show strongly increased K4me3 and strongly decreased K27me3 in *Dnmt1*$^{-/-}$ MEFs but are not changed in expression status. This suggests two points that are important for this chapter; 1) large changes in these histone modifications are not sufficient to cause gene expression changes, and, 2) the observed changes in histone modifications are not caused by changes in gene transcription. The first point highlights the importance of factors other than these histone modifications in gene expression. It could be reasoned, for example, that efficient expression of a gene requires transcription factors that may not be present in the cell tested. In this respect, it is noteworthy that genes that lose K27me3 and gain K4me3 but are not changed in expression are enriched for functions in neural development. Fibroblasts would not be expected to contain many of the transcription factors necessary for neural development as a large number of them would be expected to be neural-specific. This could explain why these genes are not changed in expression but others are. Another possible explanation could be the persistence of other epigenetic marks at these genes that are capable of maintaining repression on their own. Many PRC2 marked genes reportedly are bound by an initiated form of RNA polymerase which is suggested to be restrained somehow from proper elongation by PcG proteins (Stock *et al.*, 2007). It would be interesting to study the binding of differentially phosphorylated forms of RNA PolII at genes that lose K27me3 in *Dnmt1*$^{-/-}$ MEFs. If RNA PolII is indeed engaged at these genes then it suggests that K27me3 loss is not sufficient for polymerase release. The second point is important in relation to the mechanisms causing histone modification changes in *Dnmt1*$^{-/-}$ MEFs. It suggests that these changes are not secondary to increased gene transcription but occur independently.

Chapter 5 - Studying the role of DNA methylation in *Hox* gene regulation

5.1. Introduction

The data presented in chapter 3, together with published work, links DNA methylation to the regulation of *Hox* genes and other targets of the PRC2 complex, supporting a functional link between the two epigenetic mechanisms (Velasco *et al.,* 2010; Tao *et al.,* 2010; Tao *et al.,* 2011). Although a handful of studies have suggested that an intact DNA methylation system is required for the repression of certain PRC2-bound genes, the mechanistic link is unclear. A better comprehension of this interplay is integral to our understanding of these two mechanisms and how they influence tissue-specific gene regulation. In this chapter I aim to better understand the role that the DNA methylation system plays in repression of PRC2-bound genes. To do this I focus on the *Hox* gene clusters which are frequently studied as PRC2 targets and were shown in chapter 3 to be mis-regulated in a DNA hypomethylated cell model.

In chapter 4 I demonstrated that the PRC2 signature histone mark, K27me3, is redistributed in DNA hypomethylated cells, suggesting that the DNA methylation mark is involved in targeting of the PRC2 complex to specific regions of chromatin. The K27me3 mark at *Hox* clusters is severely reduced in the hypomethylated *Dnmt1*$^{-/-}$ MEFs despite a global increase in this mark in these cells. As the PRC2 complex is thought to be required for *Hox* gene repression, the loss of PRC2-mediated repression of *Hox* genes in *Dnmt1*$^{-/-}$ MEFs could at least partially explain their observed mis-expression in this context. However, it is not clear how DNA methylation contributes to the distribution of the K27me3 mark. It is also not clear why the K27me3 mark at some regions, such as *Hox* genes, should be so severely lost in *Dnmt1*$^{-/-}$ MEFs. It remains possible that other factors link DNA methylation to *Hox* gene repression. I reasoned that studying *Hox* clusters in more detail in relation to these points may help to inform us about how DNA methylation contributes to the regulation of these genes and also how it influences the distribution of the K27me3 mark.

To understand the relationship between DNA methylation, K27me3 and *Hox* gene regulation, I felt it important to characterise DNA methylation patterns at *Hox* clusters in more detail. Generally, our knowledge of DNA methylation patterns at *Hox* clusters is limited. Evidence suggests a complex tissue-specific and cluster-specific pattern of DNA methylation exists at *Hox* clusters. *Hox* clusters contain many CpG islands, both promoter associated and orphan (Illingworth *et al.,* 2008). Around gene promoter regions, DNA methylation at CpG islands appears to be low in most tissues but may show some tissue-specific variation (Illingworth *et al.,* 2008; Xi *et al.,* 2007; Tao *et al.,* 2010). Some *Hox*

cluster orphan CpG islands are methylated in a tissue-specific manner, but the function of these elements and their methylation status is unclear (Illingworth *et al.*, 2008). Also, there appears to differences in DNA methylation between *Hox* clusters in certain tissues, indicating that they are not always subject to the same regulatory processes (Avraham *et al.*, 2010). In multiple cancers, *Hox* gene promoters are reportedly targets for aberrant *de novo* DNA methylation (Rauch *et al.*, 2007; Avraham *et al.*, 2010). Overall, the DNA methylation pattern at *Hox* clusters is complex and not well understood. In this chapter, I first set out to map DNA methylation at *Hox* clusters in MEFs. I also aim to study some of the putative PRC2 targeting mechanisms that were described in chapter 1 (Fig. 1.7), in order to identify potential links between these processes and DNA methylation.

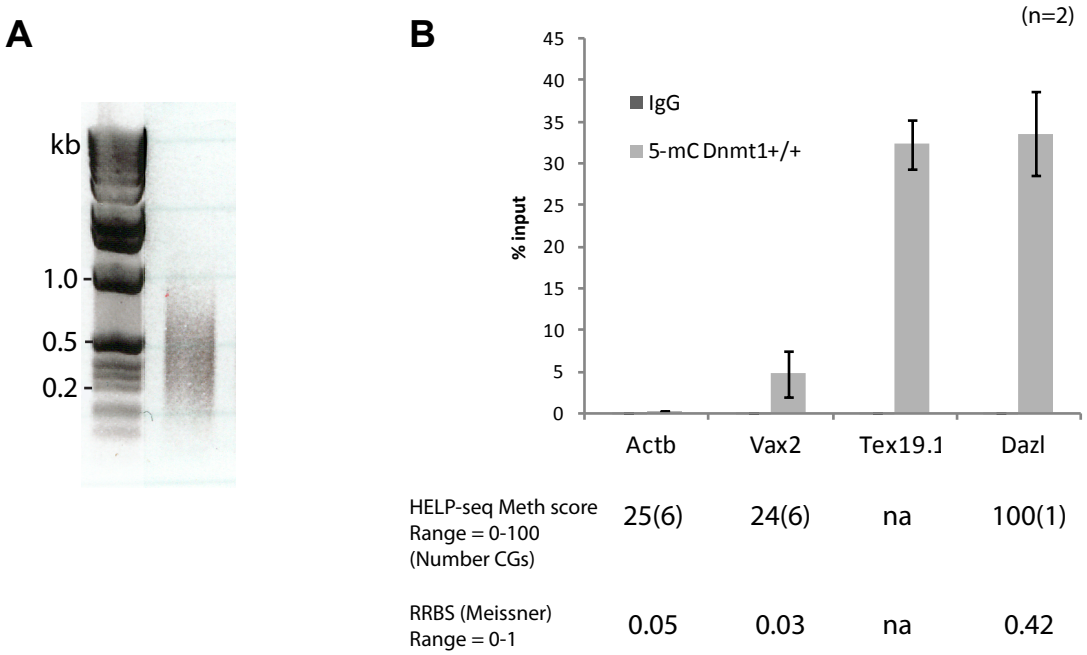
5.2. Results

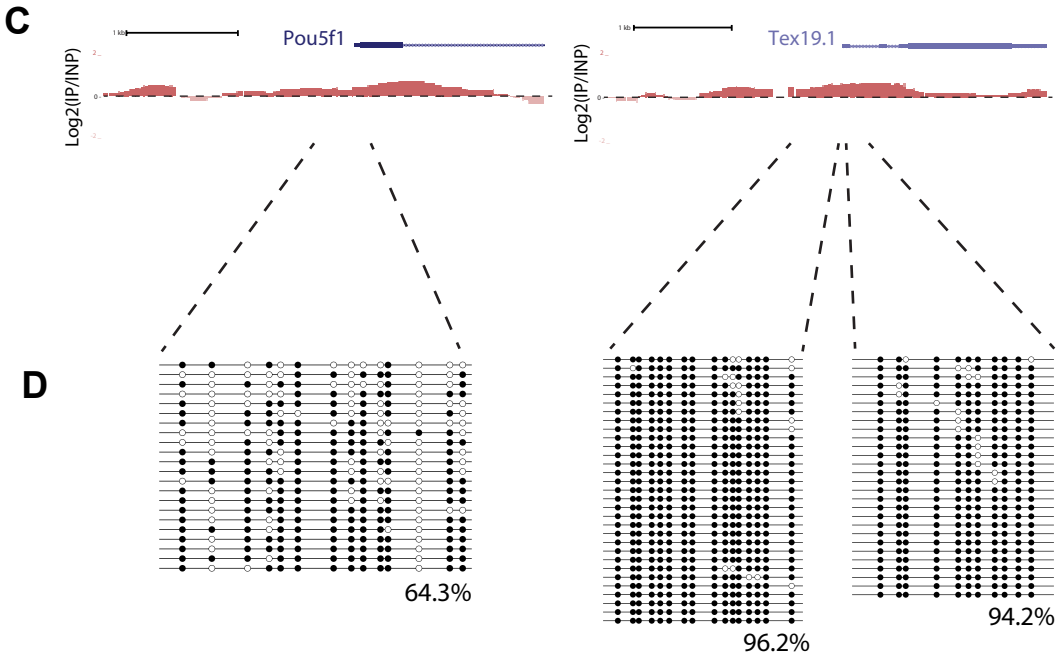
5.2.1. Characterising DNA methylation patterns at Hox *clusters in Dnmt1$^{+/+}$ and Dnmt1$^{-/-}$ MEFs*

In order to characterise the distribution of DNA methylation at *Hox* clusters in MEFs I first performed methyl-DNA-immunoprecipitation (MeDIP). MeDIP involves immunoprecipitation of DNA with an antibody that recognises 5mCpG and has been successfully used to map DNA methylation patterns on a large scale when performed in combination with microarray (MeDIP-chip) or next generation sequencing (MeDIP-seq)(for example, Borgel *et al.*, 2010). I planned to analyse immunoprecipitated DNA using a custom tiled microarray covering all four entire *Hox* clusters allowing the interrogation of intergenic and intragenic regions. Advantages of the MeDIP-chip technique include its ability to differentiate between 5mC and 5hmC, and its low cost in relation to some other DNA methylation mapping techniques. Disadvantages of this technique include the strong effect of sequence biases on signal which can make interpretation sometimes less intuitive (Pelizzola *et al.*, 2008; Down *et al.*, 2008; reviewed in Laird, 2010). Genomic DNA was sonicated to give a range of fragment lengths between ~200 and 1000bp (Fig. 5.1A). MeDIP-qPCR for control regions was performed at genes where promoter DNA methylation data was available. A reduced representation bisulfite sequencing study in wildtype MEFs by Meissner *et al.* (2008) was used, together with my HELP-tag-seq analysis on *Dnmt1*$^{+/+}$ MEFs, to pick gene promoters at each end of the methylation spectrum. *Actb* and *Vax2* promoters are each associated with CpG islands and have low level DNA methylation in MEFs in these datasets. The *Dazl* promoter is associated with a CpG island and is highly methylated in MEFs according to this data. The *Tex19.1* promoter is CpG-rich, but not associated with a CpG island (by UCSC sequence based classification), and is not included in either of these datasets but is known to be highly methylated in MEFs (shown by bisulfite sequencing in chapter 3). These promoter regions were assayed by MeDIP-qPCR and showed the expected pattern of enrichment, suggesting that the MeDIP procedure used here can successfully differentiate between regions with different methylation levels, at least at each extreme (Fig. 5.1B). Next, MeDIP enriched and input DNA was analysed in triplicate using a high-probe-density custom-designed Nimblegen tiling microarray representing all four *Hox* clusters and other selected genomic regions (the same microarray used in chapter 3

Figure 5.1 - MeDIP-chip can be used to approximate DNA methylation levels

A. Sonicated DNA from *Dnmt1*$^{+/+}$ MEFs sized on an agarose gel. B. MeDIP-qPCR for control gene promoters in *Dnmt1*$^{+/+}$ MEFs. Error bars represent +/- standard error of the mean. The methylation level measured by HELP-seq and RRBS (from Meissner *et al.,* 2008) are shown below graph. C. UCSC browser image showing MeDIP-chip signal (log2(IP/INP)) in *Dnmt1*$^{+/+}$ MEFs. D. Bisulfite sequencing for indicated regions of the *Pou5f1* and *Tex19.1* promoters. Filled and open circles indicate methylated and unmethylated CpG dinucleotides respectively. The percentage of CpGs that are methylated in each region is shown below each image.





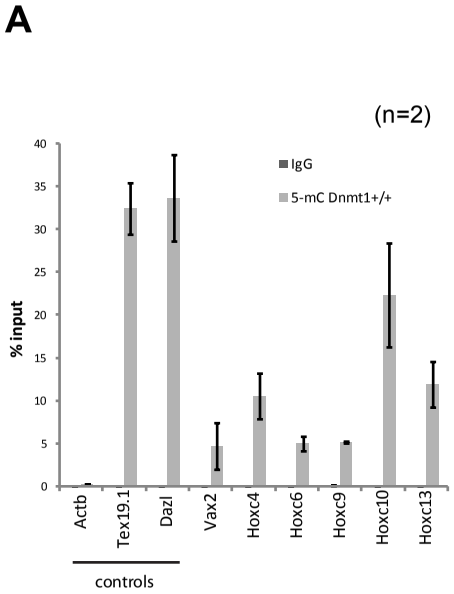
to analyse cDNA signal). For experimental setup, microarray details and data normalisation techniques employed see chapter 2. The microarray signal at methylated regions where I have previously performed bisulfite sequencing was assessed as control (Fig. 5.1C). Both the *Pou5f1* and *Tex19.1* promoter showed an enrichment in MeDIP-chip signal as expected (Fig. 5.1C). The DNA methylation status of these regions by bisulfite sequencing in *Dnmt1*$^{+/+}$ MEFs is also shown (Fig. 5.1D). I did not perform MeDIP on *Dnmt1*$^{-/-}$ MEF DNA as preliminary experiments indicated that this technique does not allow inter-sample comparison of methylation patterns where one sample is severely hypomethylated relative to another (data not shown).

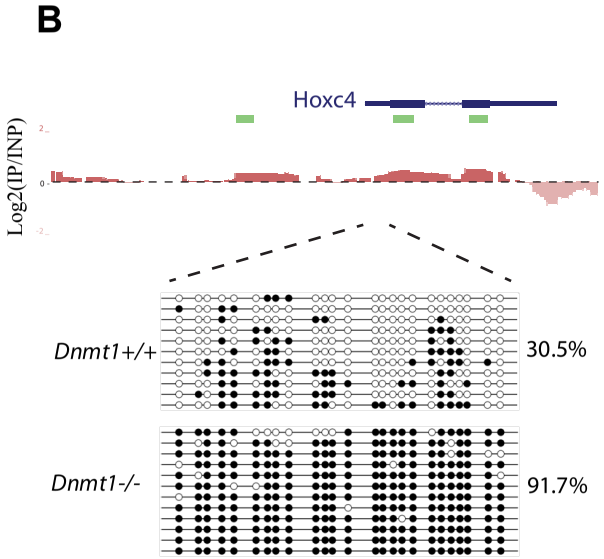
I next investigated the DNA methylation status at promoters of *Hox* genes. I measured the signal in immunoprecipitated DNA as a percentage of input using MeDIP-qPCR of *HoxC* promoters in *Dnmt1*$^{+/+}$ MEFs (Fig. 5.2A). Each of the analysed promoters showed an enrichment between that of the *Actb* and *Dazl* promoters. The *Hoxc10* promoter showed the highest signal by MeDIP-qPCR of the analysed *HoxC* genes (Fig. 5.2A). By MeDIP-chip in *Dnmt1*$^{+/+}$ MEFs most *Hox* promoters showed a peak of signal that encompassed the transcription start site (Fig. 5.2B to G upper). However, peaks were observed at all CpG-rich regions within *Hox* clusters (Fig. 5.3). I performed bisulfite sequencing of selected promoter regions of *HoxC* and *HoxA* genes in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs in order to confirm the observations made by MeDIP-chip and also to quantify DNA methylation in both cell lines. I felt this was necessary as MeDIP is very sensitive to CpG density and so cannot by itself be used quantitatively (Pelizzola *et al.*, 2008; Down *et al.*, 2008; reviewed in Laird, 2010). In *Dnmt1*$^{+/+}$ MEFs the regions analysed showed between 19% and 55% overall methylation (Fig. 5.2B to G lower). This suggests that the patterns observed by MeDIP-chip are highly influenced by local CpG-content as even low to intermediate DNA methylation content appears as a peak if present in CpG-rich regions. In *Dnmt1*$^{-/-}$ MEFs, every *Hox* promoter region, apart from *Hoxc4*, showed the expected hypomethylation by bisulfite sequencing (Fig. 5.2B to G lower). Bizarrely, the *Hoxc4* promoter is hypermethylated in *Dnmt1*$^{-/-}$ MEFs, an observation that was repeated to ensure it was not an experimental error (Fig. 5.2B lower). These results suggest that in *Dnmt1*$^{+/+}$ MEFs, the *Hox* gene promoters analysed contain an intermediate level of DNA methylation, and, at most regions, the vast majority of this DNA methylation is lost in *Dnmt1*$^{-/-}$ MEFs.

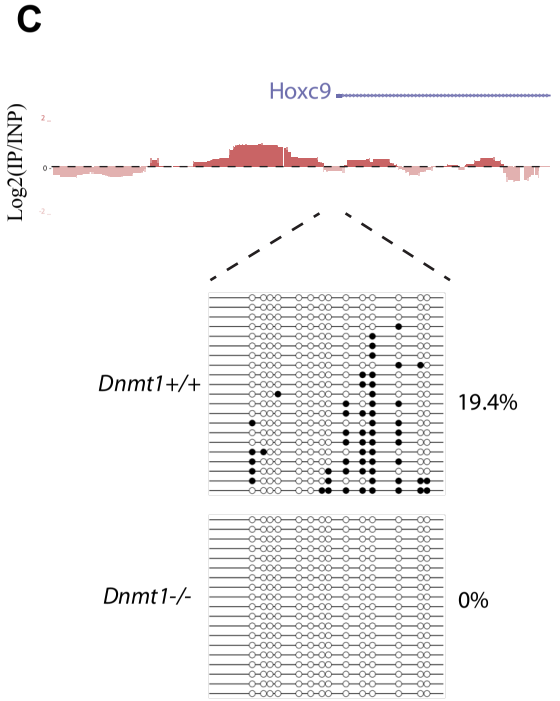
I next analysed DNA methylation at intergenic regions of *Hox* clusters, focussing on orphan CpG islands (CpG islands not associated with the 5' end of annotated genes). Some orphan CpG islands within *Hox* clusters have been shown to be methylated in a tissue-specific manner (Illingworth *et al.*, 2008). I used two definitions of CpG islands to

Figure 5.2 - DNA methylation at *Hox* gene promoters in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs

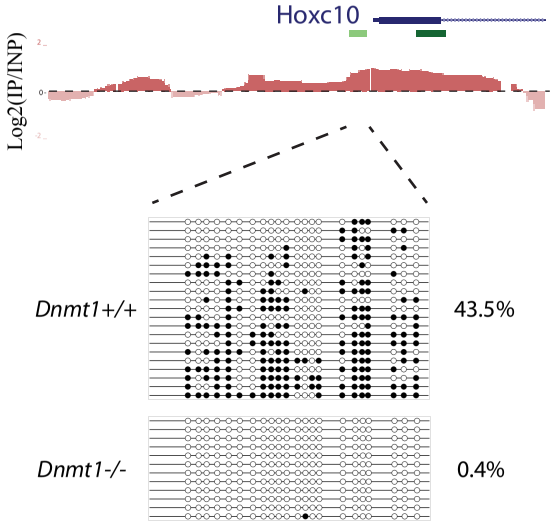
A. MeDIP-qPCR for control and *Hox* gene promoters in *Dnmt1*$^{+/+}$ MEFs. Error bars indicate standard error of the mean. B-G. UCSC browser image (top panel) showing MeDIP-chip signal for *Dnmt1*$^{+/+}$ MEFs. Position of Refseq (mm9) genes (blue) and UCSC CpG islands (green) are shown. Bisulfite sequencing of the indicated regions in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs is shown below. Filled and open circles represent methylated and unmethylated CpGs respectively. The overall percentage of methylation is shown.



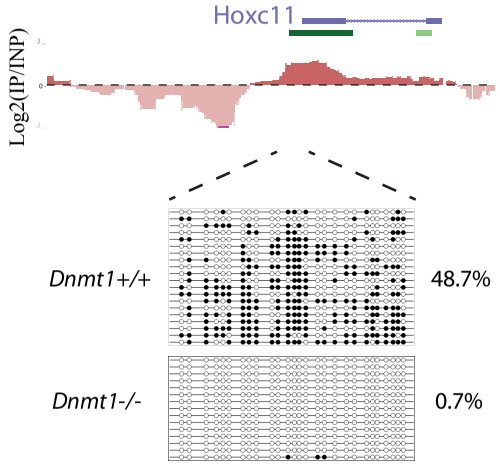




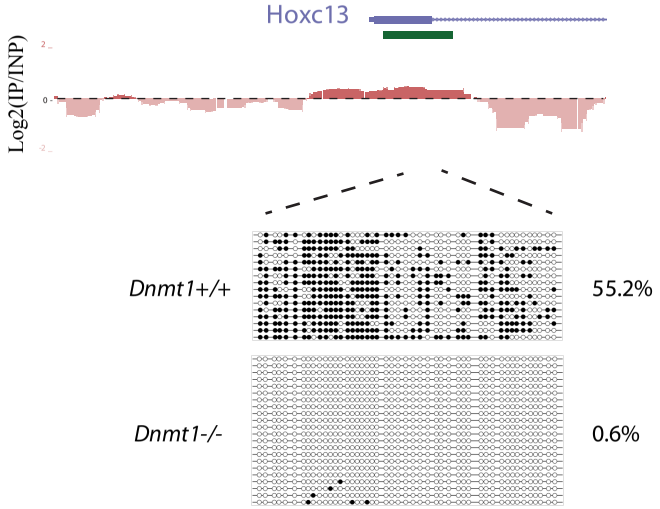
D


Hoxc10
Log2(IP/INP)
Dnmt1+/+
43.5%
Dnmt1-/-
0.4%

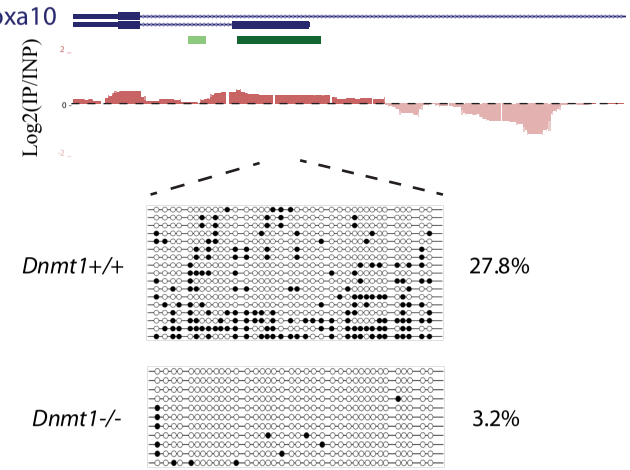
E


Hoxc11
Log2(IP/INP)
Dnmt1+/+
48.7%
Dnmt1-/-
0.7%

F


Hoxc13
Log2(IP/INP)
Dnmt1+/+
55.2%
Dnmt1-/-
0.6%

G


Hoxa10
Log2(IP/INP)
Dnmt1+/+
27.8%
Dnmt1-/-
3.2%

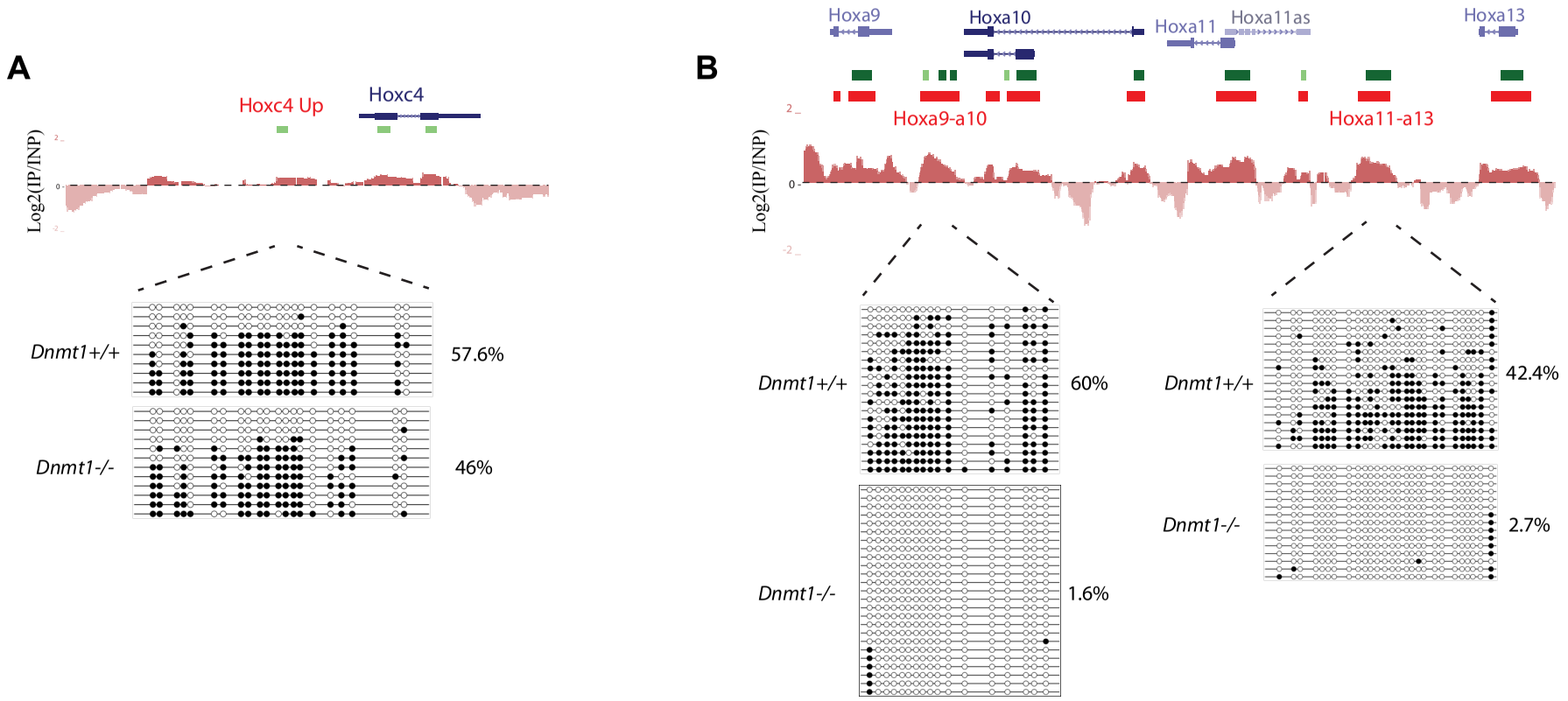
determine their locations in *Hox* clusters. The first was from the UCSC browser and is based on DNA sequence characteristics. The second is from Illingworth *et al.,* (2008 and 2010) and is based on enrichment in at least one tissue studied using a CXXC-domain enrichment technique. I included the latter CpG island list as the UCSC method is thought to exclude many regions that show characteristics of annotated CpG islands in the mouse genome (Illingworth *et al.,* 2008; 2010). By MeDIP-chip in *Dnmt1*$^{+/+}$ MEFs, most orphan CpG islands defined using the two methods showed enriched signal relative to flanking regions (Fig. 5.3A to C upper). As MeDIP signal is strongly affected by CpG content, I also assayed DNA methylation at many orphan CpG islands by bisulfite sequencing. These regions showed between 42% and 60% overall methylation in *Dnmt1*$^{+/+}$ MEFs confirming that they contain considerable DNA methylation in these cells (Fig. 5.3A to C lower). In *Dnmt1*$^{-/-}$ MEFs, DNA methylation was much reduced at these regions, apart from two, *Hoxc4 Up* and *Hoxc13-c12*, which did not show decreased methylation (Fig. 5.3A to C lower). These results suggest that intergenic CpG islands within *HoxC* and *HoxA* clusters contain an intermediate level of DNA methylation in *Dnmt1*$^{+/+}$ MEFs.

The *HoxA*, *HoxC* and *HoxD* clusters show active and inactive portions in *Dnmt1*$^{+/+}$ MEFs in regard to gene expression (shown in chapter 3). If local DNA methylation patterns are important for this expression pattern then one might expect differences in DNA methylation between the two parts of the clusters. I investigated this possibility at the *HoxA* cluster by MeDIP-chip. I used the cDNA custom tiling array data from chapter 3 to arbitrarily split the *HoxA* cluster into two parts, an 'active' *Hoxa1-a7* portion and an 'inactive' *Hoxa9-13* portion (Fig. 5.4A). Plotting the cDNA signal at exonic probes for the two portions of the cluster confirms the difference in activity (Fig. 5.4B, $p<2.2e-16$ by Mann-Whitney U test). By eye, no gross difference in MeDIP-chip signal are observed between the two portions of the cluster (Fig. 5.4A). This is confirmed by plotting the MeDIP-chip signal for every probe in each portion, and for probes covering the promoter region of each gene (Fig. 5.4C and D, $p>0.05$ by Mann Whitney U test). This suggests that the inactive portion of the *HoxA* cluster in *Dnmt1*$^{+/+}$ MEFs is not more DNA methylated than the active portion, and that an overall difference in DNA methylation content does not influence the partitioned expression of this cluster.

5.2.2. Low levels of DNA methylation at Hox CpG islands in primary MEFs suggests culture-induced hypermethylation in Dnmt1$^{+/+}$ *MEFs*

Figure 5.3 - DNA methylation at orphan CpG islands within *Hox* clusters in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs

A-C. UCSC browser image (upper panel) showing MeDIP-chip data on *Dnmt1*$^{+/+}$ MEFs. Position of Refseq genes (blue), UCSC CpG islands (green) and Illingworth CpG islands (red)(see Methods) are shown. The name given to each intergenic CpG island is shown (red). Bisulfite sequencing (lower panel) for indicated regions in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Percentage of CpGs that are methylated is shown. Filled and open circles represent methylated and unmethylated CpGs respectively.



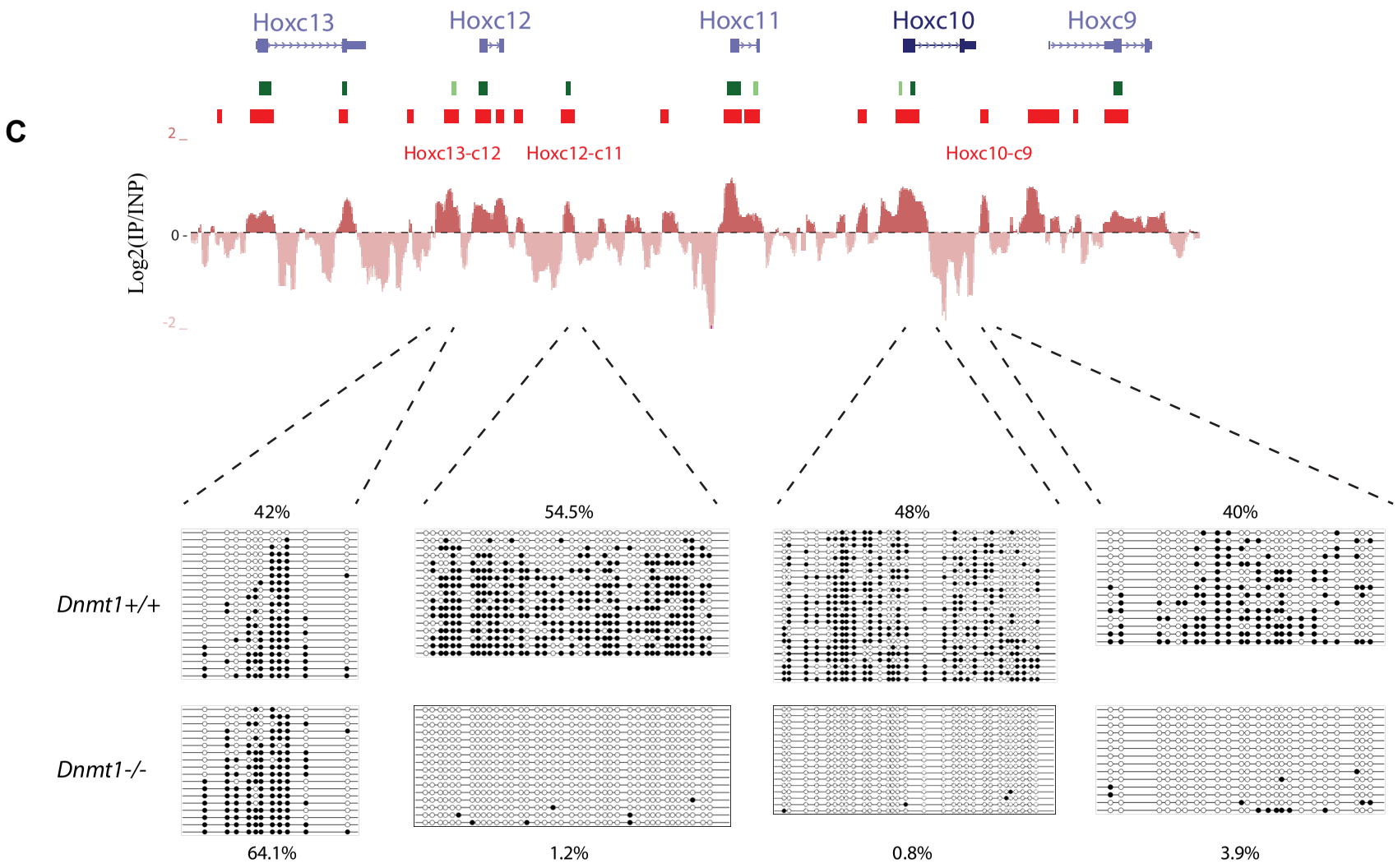
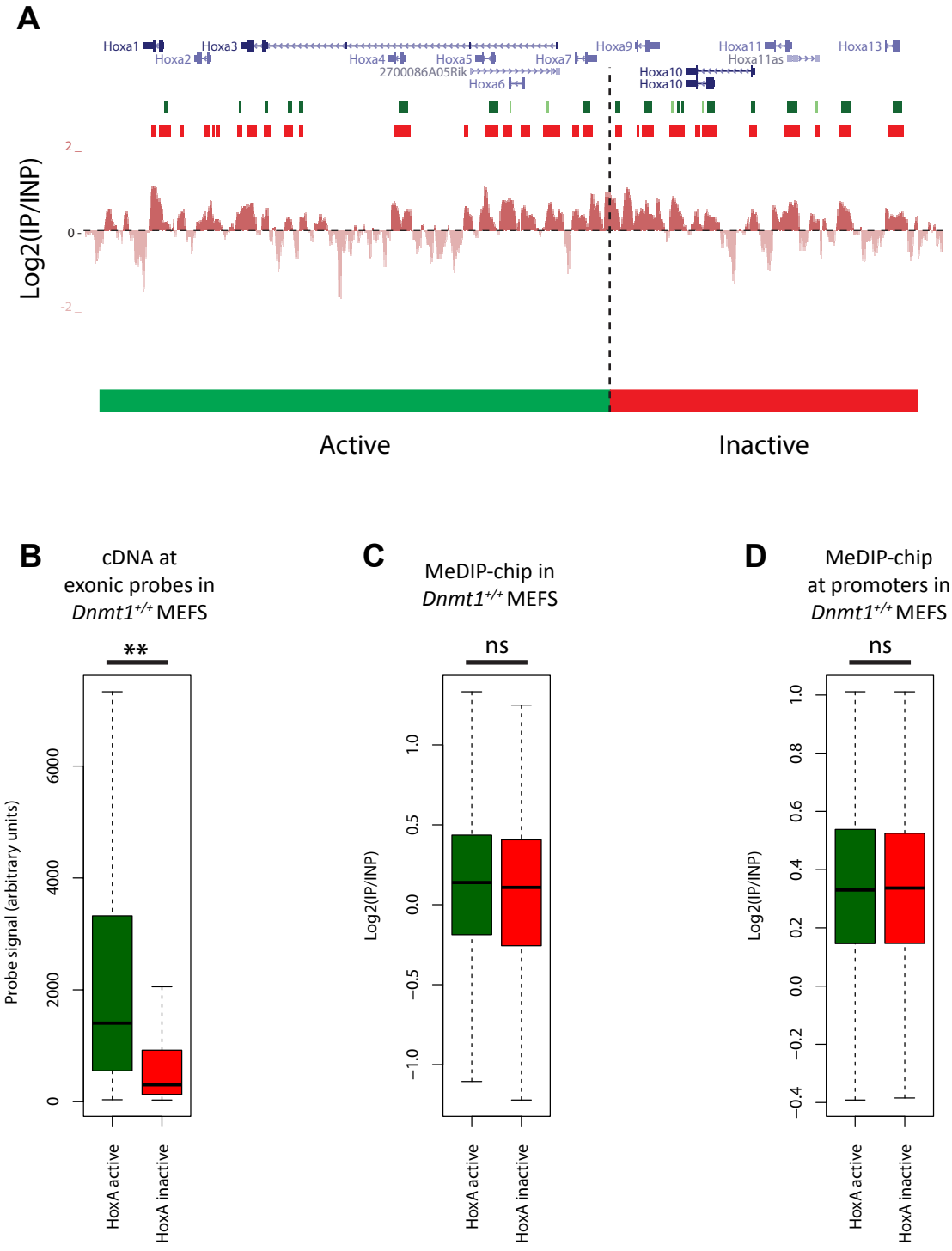

C
Hoxc13
Hoxc12
Hoxc11
Hoxc10
Hoxc9
2
Hoxc13-c12
Hoxc12-c11
Hoxc10-c9
Log2(IP/INP)
0
-2
42%
54.5%
48%
40%
Dnmt1+/+
Dnmt1-/-
64.1%
1.2%
0.8%
3.9%

Figure 5.4 - The inactive portion of the *HoxA* cluster shows no enrichment for DNA methylation in *Dnmt1*$^{+/+}$ MEFs

A. UCSC browser image showing MeDIP data from tiling array for *HoxA* cluster. Location of RefSeq genes (blue), UCSC CpG islands (green rectangles) and Illingworth CpG islands (red rectangles)(see Methods) are shown. Partition of cluster into 'active' and 'inactive' portions based on cDNA data is shown below (red and green bars). B to D. Comparison of probe signals from the 'active' and 'inactive' portions of the cluster for cDNA at exonic probes (B) DNA methylation by MeDIP (C) and DNA methylation by MeDIP at promoter regions only (D). ** and 'ns' indicate p<2.2e-16 and p>0.1 respectively by Mann Whitney-U test between the two groups.



DNA methylation patterns have been suggested to change in cell culture and during aging, particularly increasing at silent developmental genes and those marked by K27me3 (Meissner *et al.*, 2008; Borgel *et al.*, 2010; Rakyan *et al.*, 2010). *Hox* genes also appear to be susceptible to hypermethylation in a variety of cancers (Rauch *et al.*, 2007; Avraham *et al.*, 2010). I thought it possible that some of the DNA methylation observed in figures 5.2 and 5.3 at *Hox* CpG islands has been acquired in cell culture. I felt that this was possible as an intermediate DNA methylation content of CpG islands, as observed in figures 5.2 and 5.3, is thought to be rare in normal cells, as the vast majority of CpGs are thought to show a bimodal distribution of DNA methylation (Meissner *et al.*, 2008; Lister *et al.*, 2009). To address this possibility I isolated genomic DNA from primary (passage 1) MEFs and performed bisulfite sequencing of many of the same loci as analysed in *Dnmt1*$^{+/+}$ MEFs (Fig. 5.5). I included both promoter regions and intergenic CpG islands in this analysis. All of the regions analysed, apart from the *Hoxc9* promoter, show lower levels of DNA methylation in primary MEFs than in the cultured *Dnmt1*$^{+/+}$ MEFs (Table 5.1). Further DNA methylation analysis will need to be performed to understand this difference. This data suggest that most of the DNA methylation observed in *Dnmt1*$^{+/+}$ MEFs was accrued in cell culture and is not representative of their tissue of origin.

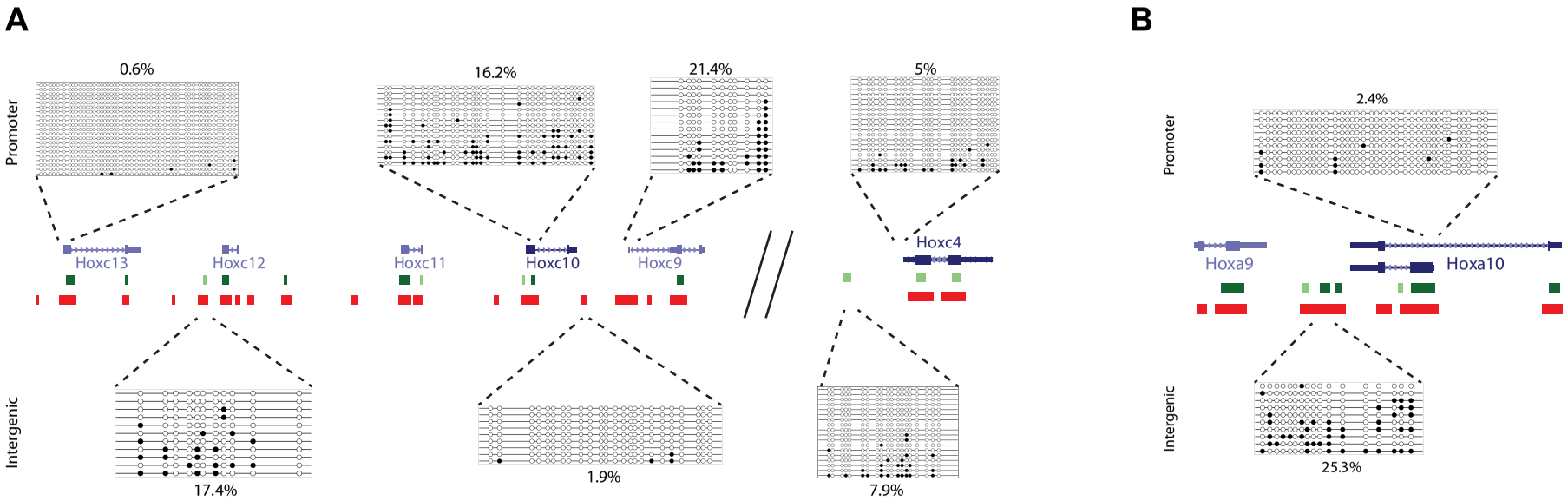
Region	Location	DNA methylation in *Dnmt1*$^{+/+}$ MEFs (%) (cultured)	DNA methylation in p1 MEFs (%) (primary)
Hoxc13	Promoter	55.2	0.6
Hoxc10	Promoter	43.5	16.2
Hoxc9	Promoter	19.4	21.4
Hoxc4	Promoter	30.5	5
Hoxa10	Promoter	27.8	2.4
Hoxc13-c12	Intergenic	42	17.4
Hoxc10-c9	Intergenic	40	1.9
Hoxc4 Up	Intergenic	57.6	7.9
Hoxa9-a10	Intergenic	60	25.3

Table 5.1 - DNA methylation as measured by bisulfite sequencing in cultured *Dnmt1*$^{+/+}$ MEFs and primary MEFs at selected regions. Regions with lower methylation in primary MEFs than *Dnmt1*$^{+/+}$ MEFs are coloured red. The location of the intergenic CpG island regions analysed are shown in figure 5.3.

In conclusion to this part of the chapter, despite efforts to characterise DNA methylation patterns at *Hox* genes in MEFs, data presented here suggest that *Dnmt1*$^{+/+}$ MEFs have

Figure 5.5 - DNA methylation at *Hox* clusters in primary MEFs

A. Bisulfite sequencing results for promoter and intergenic regions in the *HoxC* cluster in primary anterior trunk MEFs. Open and filled circles represent unmethylated and methylated CpGs respectively. The location of regions relative to RefSeq genes (blue) , UCSC CpG islands (green) and Illingworth CpG islands (red) is shown. The percentage of methylated CpGs is shown. B. As 'A' but for regions within the *HoxA* cluster.



accumulated DNA methylation in culture and so may not represent a good model to study DNA methylation patterns. I suggest that in primary MEFs, promoter associated and orphan CpG islands within *HoxA* and *HoxC* clusters are hypomethylated. Further investigation is required to determine the pattern of DNA methylation within *Hox* clusters in normal MEFs and primary tissues, and to connect this with the regulation of these gene clusters.

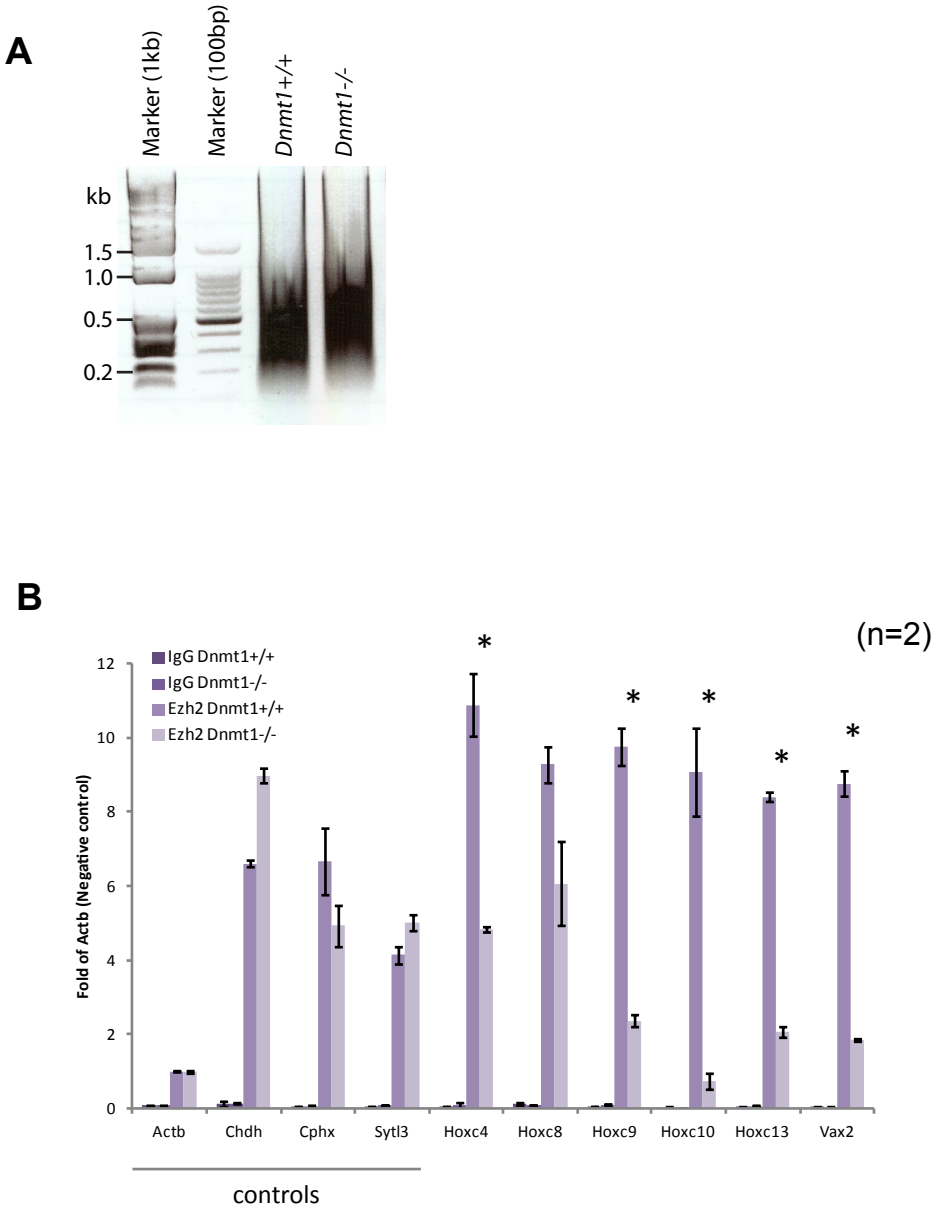
5.2.3. The association of PRC2 component Ezh2 with HoxC genes is attenuated in Dnmt1$^{-/-}$ MEFs

In order to begin to investigate the mechanism of K27me3 loss at *Hox* genes in DNA hypomethylated MEFs, I first investigated the binding of Ezh2 to chromatin at target genes. Ezh2 is a core and catalytic subunit of the PRC2 complex and remains bound to PRC2 repressed chromatin (Bracken *et al.*, 2006). I reasoned that lack of K27me3 at *Hox* genes in *Dnmt1*$^{-/-}$ MEFs could be due to; 1) impaired PRC2 binding to these regions, or, 2) reduced PRC2 catalytic activity at these regions despite no change in PRC2 binding. To differentiate between these two possibilities I performed formaldehyde cross-linking of cells followed by chromatin immunoprecipitation (X-ChIP) for Ezh2. Chromatin was sonicated to give fragments approximately 200 - 1000bp in length (Fig. 5.6A). X-ChIP-qPCR showed low enrichment for Ezh2 at the negative control *Actb* promoter compared with positive control *Hox* genes in *Dnmt1*$^{+/+}$ MEFs suggesting that detection of Ezh2 binding using this technique is specific (Fig. 5.6B). Three gene promoters that show no decrease in K27me3 in *Dnmt1*$^{-/-}$ MEFs also show no decrease in Ezh2 binding (Fig. 5.6B). All of the *HoxC* and *Vax2* genes analysed showed decreased association with Ezh2 in *Dnmt1*$^{-/-}$ MEFs, and five were significantly lower by two-tailed Student's t-test (Fig. 5.6B, $p<0.05$). This result indicates that association of the PRC2 complex with *Hox* promoter chromatin is attenuated in *Dnmt1*$^{-/-}$ MEFs. The enrichment of Ezh2 at many of the assayed genes in *Dnmt1*$^{-/-}$ MEFs approaches that of the negative control *Actb* promoter suggesting a strong reduction in binding (Fig. 5.6B). This is consistent with the observed loss of the K27me3 mark at these regions and suggests that K27me3 loss may be due to loss of PRC2 complex binding to these regions of chromatin.

5.2.4. The involvement of the CTCF protein and boundary elements?

Figure 5.6 - The association of PRC2 component Ezh2 with *Hox* genes is attenuated in *Dnmt1*$^{-/-}$ MEFs

A. DNA from sonicated cross-linked chromatin was extracted and sized on an agarose gel. B. X-ChIP-qPCR for control and target gene promoters after Ezh2 immunoprecipitation from *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars indicate +/- standard error of the mean (n=2). * indicates $p<0.05$ by two-tailed Student's t-test between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs.

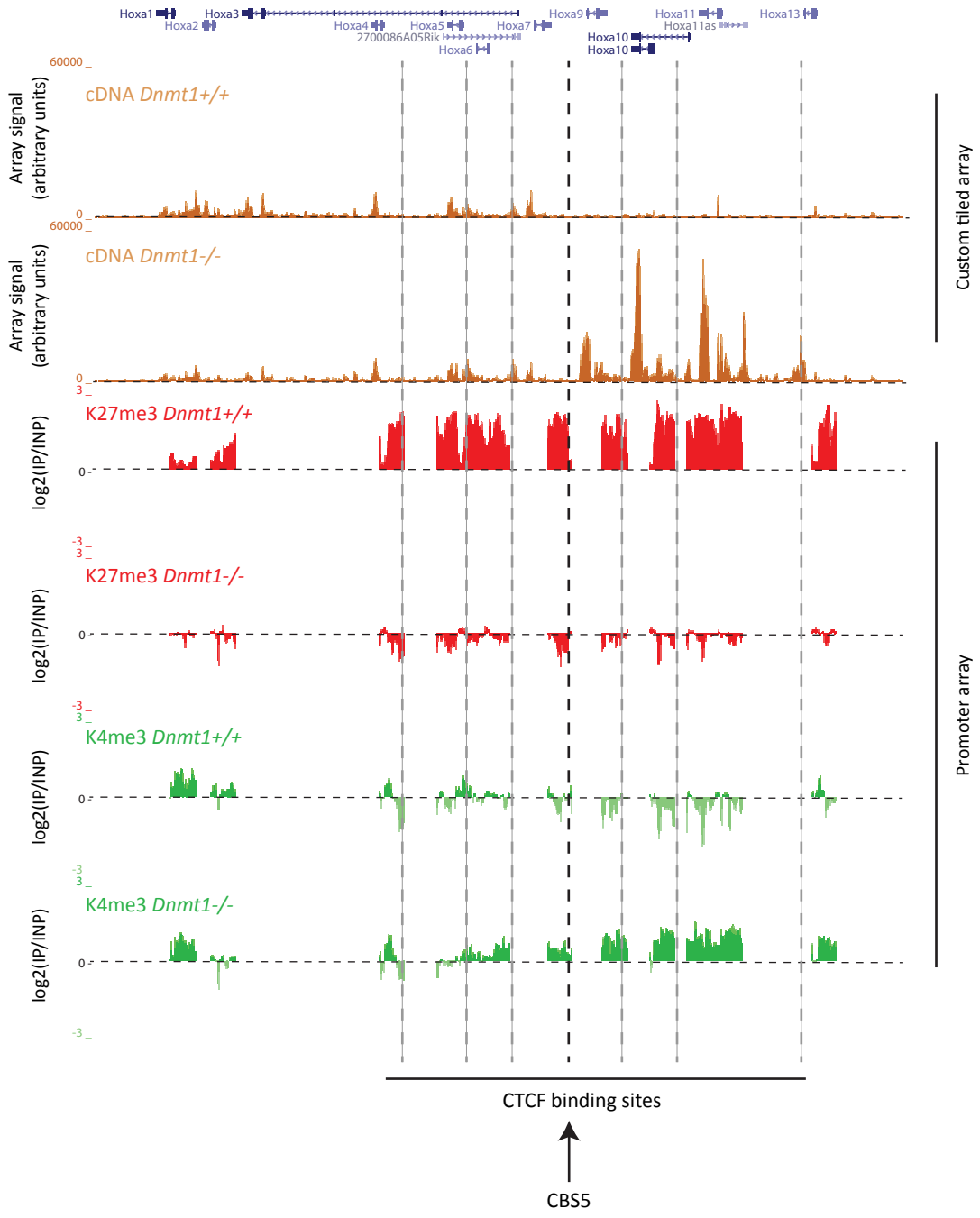


CTCF is a DNA binding protein that has been implicated in mediating chromatin contacts and forming boundary and insulator elements (Philips and Corces, 2009). The formation of intra-chromosomal loops has been linked to PRC2 binding and gene regulation, although this process is not well understood. For example, CTCF has been suggested to interact with the PRC2 component Suz12 and may recruit PRC2 to allow K27me3 to spread within a chromosomal loop, to mediate imprinting of the *Igf2* gene (Li *et al.*, 2008; Zhang *et al.*, 2011b). CTCF is bound to multiple positions in *Hox* clusters and has been suggested to partition the different domains of the clusters through its activity as an insulator (Kim *et al.*, 2011; Soshnikova *et al.*, 2010). Several factors made me want to investigate CTCF binding in my system: 1) Analysis in chapter 4 suggested that K27me3 is often lost from multiple genes in tandem in *Dnmt1*$^{-/-}$ MEFs. One explanation for this observation could be that CTCF-mediated chromatin loops are lost which may normally act to recruit/restrict K27me3. 2) CTCF depletion has been reported to result in altered K27me3 and expression of *Hox* genes in a manner that depends on their location within the cluster (Kim *et al.*, 2011; Soshnikova *et al.*, 2010). 3) The pattern of gene expression and histone mark changes at *Hox* clusters in *Dnmt1*$^{-/-}$ MEFs occurs around a particular CTCF binding site. Increased cDNA and K4me3 was observed at *Hoxa9-13* genes (Fig. 5.7). A postulated boundary element, termed CBS5, exists between *Hoxa7* and *Hoxa9* where CTCF is bound in many cell types (Kim *et al.*, 2011)(Fig. 5.7). Similar patterns of expression and histone modification changes are observed at the *HoxC* and *HoxD* clusters where a homologous CTCF binding site has also been found (Kim *et al.*, 2007; Soshnikova *et al.*, 2010). The *HoxA* CBS5 locus has been suggested to form an intra-chromosomal loop with a downstream element, acting to partition the cluster (Kim *et al.*, 2011). One possibility is that the activity of this boundary element is lost in *Dnmt1*$^{-/-}$ MEFs contributing to the observed changes. To address this hypothesis I examined CTCF binding to this element by X-ChIP. I used primers for control regions that were expected to be constitutively bound by CTCF (*c-Myc* 5' and *Pax6* 6a) or not bound by CTCF (*c-Myc* -5kb) to ensure that the assay was effective (A. Buckle, personal communication)(Fig. 5.7B and C). Large enrichment for CTCF binding was observed at both the *HoxA* CBS5 and *HoxC* c8-c9 loci in *Dnmt1*$^{+/+}$ MEFs, relative to the positive controls, suggesting that CTCF is bound to both of these loci in these cells. Reduced CTCF binding was observed in *Dnmt1*$^{-/-}$ MEFs at the *HoxA* CBS5 and *HoxC* c8-c9 loci, while no decreases were observed at the control genes (Fig. 5.7C). This experiment was performed only once and will need to be repeated before conclusions are drawn. However, this suggests that CTCF binding to *HoxA* and *HoxC* boundary elements is reduced in *Dnmt1*$^{-/-}$

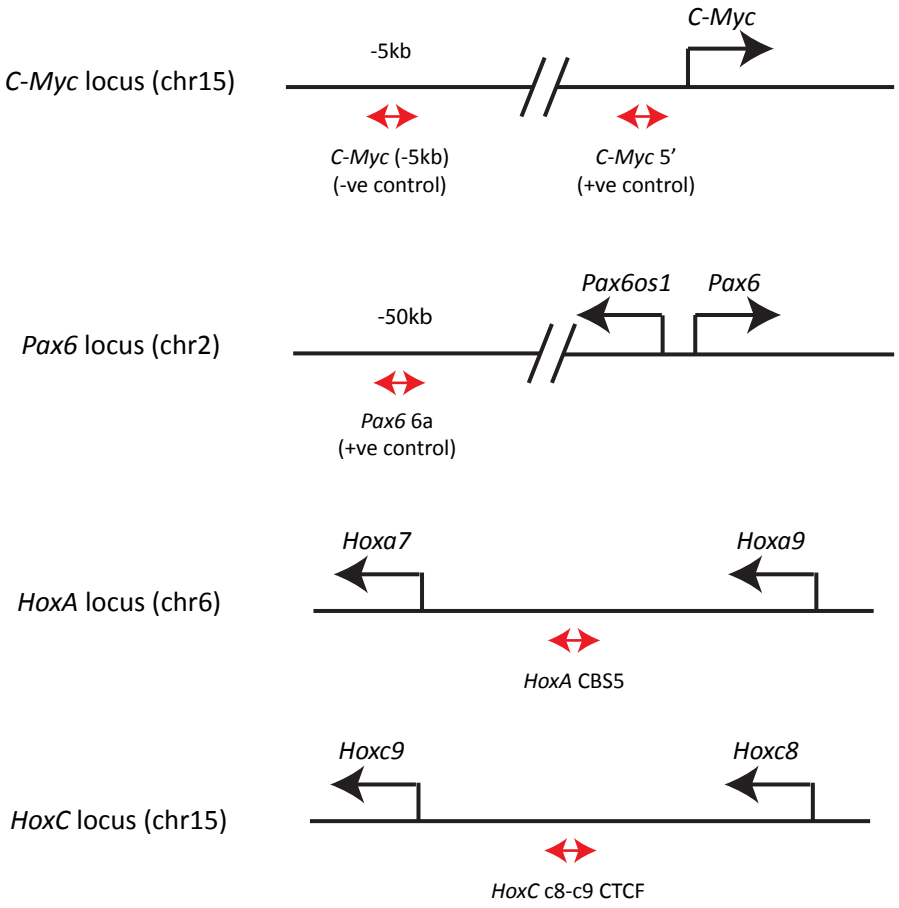
Figure 5.7 - Broad histone modification and transcriptional changes occur around a proposed boundary element in *Hox* clusters

A. UCSC browser images of cDNA-chip (custom tiled array) and N-ChIP-chip (promoter array) data at the *HoxA* cluster in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Position of Refseq genes (blue) are shown. Locations of putative CTCF binding sites are shown as grey and black dashed lines. The location of the CBS5 boundary element is shown as a black dashed line (Kim *et al.*, 2011). B. Primer binding locations relative to the transcription start site of indicated genes. The name of the primer pair is shown below the location of the amplicon (red arrows). C. X-ChIP-qPCR for CTCF in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars indicate +/- standard error of qPCR (n=1).

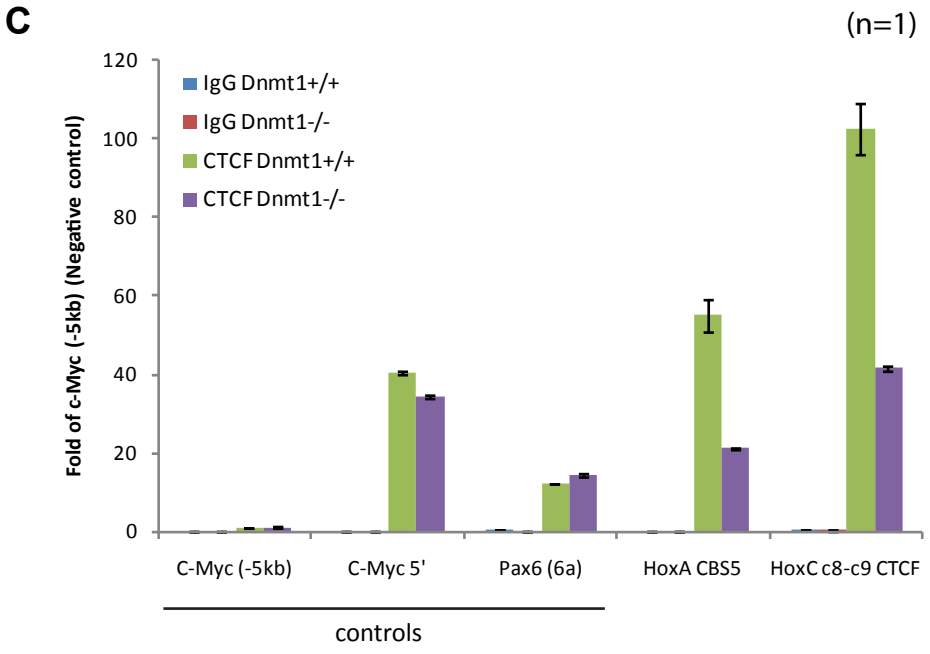
A



B


C-Myc locus (chr15)
-5kb
C-Myc
C-Myc (-5kb)
(-ve control)
C-Myc 5'
(+ve control)
Pax6 locus (chr2)
-50kb
Pax6os1
Pax6
Pax6 6a
(+ve control)
HoxA locus (chr6)
Hoxa7
Hoxa9
HoxA CBS5
HoxC locus (chr15)
Hoxc9
Hoxc8
HoxC c8-c9 CTCF

C


(n=1)
Fold of c-Myc (-5kb) (Negative control)
120
100
80
60
40
20
0
IgG Dnmt1+/+
IgG Dnmt1-/-
CTCF Dnmt1+/+
CTCF Dnmt1-/-
C-Myc (-5kb)
C-Myc 5'
Pax6 (6a)
HoxA CBS5
HoxC c8-c9 CTCF
controls

MEFs, correlating with the spreading of K4me3 and gene expression into the normally repressed *Hox9-13* domains.

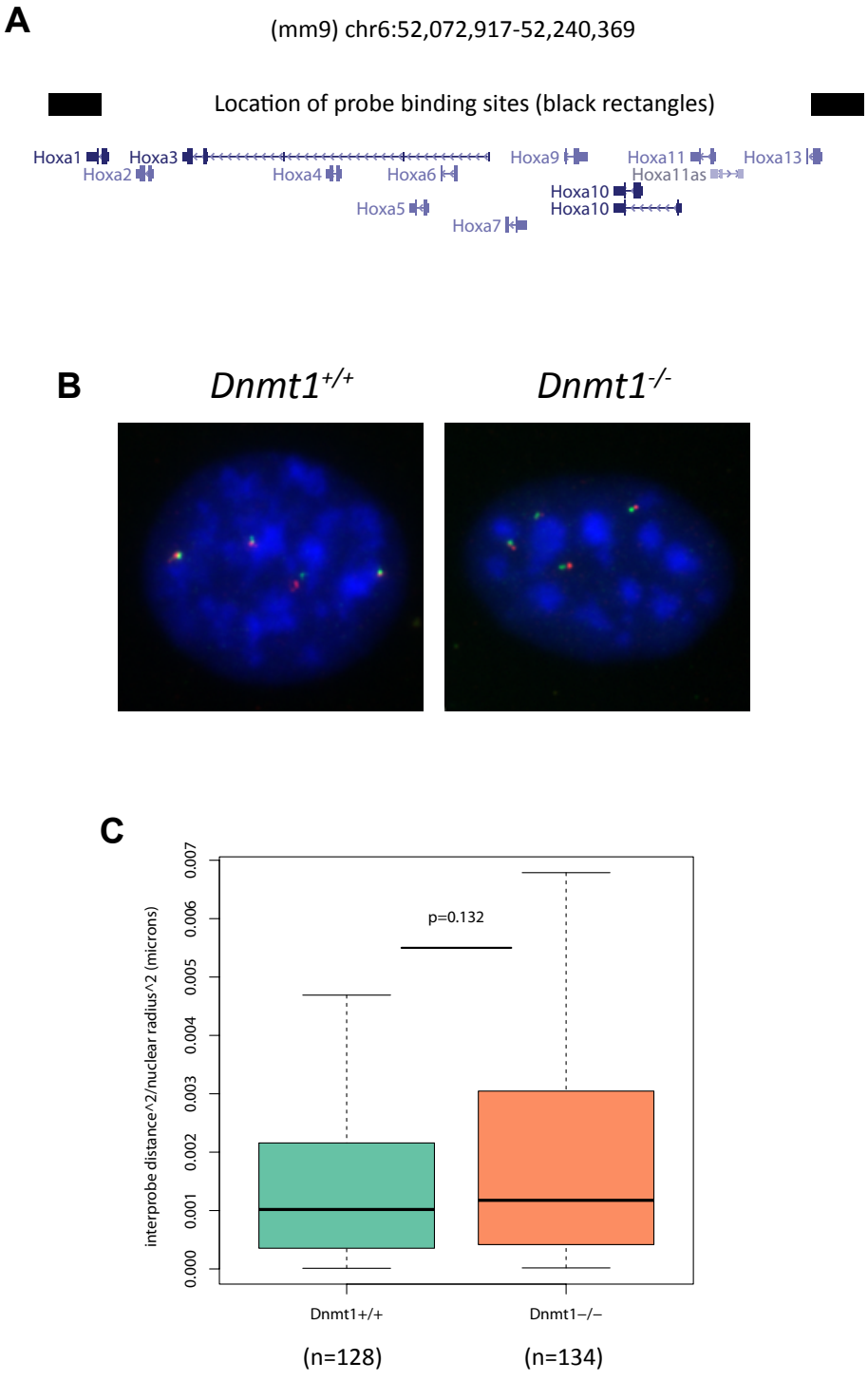
Altered CTCF binding may be expected to affect the higher order chromatin organisation of the *Hox* clusters. The higher order chromatin organisation at *Hox* clusters, as with all loci, is currently not well understood. Chromatin conformation assays have begun to shed light on this aspect of chromatin biology. *Hox* clusters appear to form multiple intra-cluster contacts in undifferentiated pluripotent cells when the entire clusters are silent (Ferraiuolo *et al.,* 2010). Upon differentiation, a change in chromatin organisation is suggested to occur, including an overall decompaction of the loci (Eskeland *et al.,* 2010). In some differentiated cells, *Hox* clusters may form a partitioned chromatin organisation with actively transcribed portions of the cluster forming an active 'chromatin hub' (Noordermeer *et al.,* 2011). Polycomb protein binding and the K27me3 mark have also been suggested to influence higher order chromatin organisation (Eskeland *et al.,* 2010; Bantignies *et al.,* 2011). In order to investigate a potential change in chromatin organisation in *Dnmt1*$^{-/-}$ MEFs, I performed two-dimensional fluorescence *in situ* hybridisation (2D FISH) using labelled probes that span the *HoxA* cluster (performed with help from S. Boyle)(Fig. 5.8A). The vast majority of *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs showed four pairs of spots indicating that they each contain four copies of this locus (e.g. Fig. 5.8B). I measured the interprobe distance of >120 spot pairs as a measure of chromatin compaction. No significant difference was observed between the two cell types (Fig. 5.8C). This suggests that at this resolution, no difference in chromatin compaction is observed in *Dnmt1*$^{-/-}$ MEFs, despite the observed changes in histone marks and gene expression. Further work is required here to determine the cause and effect of reduced CTCF binding to these sites.

5.2.5. Increased intergenic transcription and H3K36me3 are associated with K27me3 loss

Non-coding RNA (ncRNA) transcription has been suggested to inhibit PcG repression and alter nearby gene expression in certain contexts, particularly within *Hox* clusters (Cavalli and Paro, 1998; Cavalli and Paro, 1999; Bender *et al.,* 2002; Hogga *et al.,* 2002; Rank *et al.,* 2002; Schmitt *et al.,* 2005; Sanchez-Elsner *et al.,* 2006; Dinger *et al.,* 2008; Wang *et al.,* 2011; Bertani *et al.,* 2011). The precise mechanisms by which ncRNA modulates *Hox* expression is unclear. Certain ncRNA has been suggested to act by recruiting TrxG proteins to stimulate transcriptional activation of nearby genes (Sanchez-Elsner *et al.,* 2006; Wang *et*

Figure 5.8 - Chromatin compaction by FISH of the *HoxA* locus is unchanged in *Dnmt1*$^{-/-}$ MEFs

A. Probe binding locations relative to the *HoxA* cluster. B. Representative FISH images using the probes in 'A' on *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Note that both nuclei have four copies of the probe binding sites. C. Boxplot showing interprobe distance normalised to nuclear area. 'N' represents the number of spot pairs measured. P-value is from a Mann-Whitney-U test.

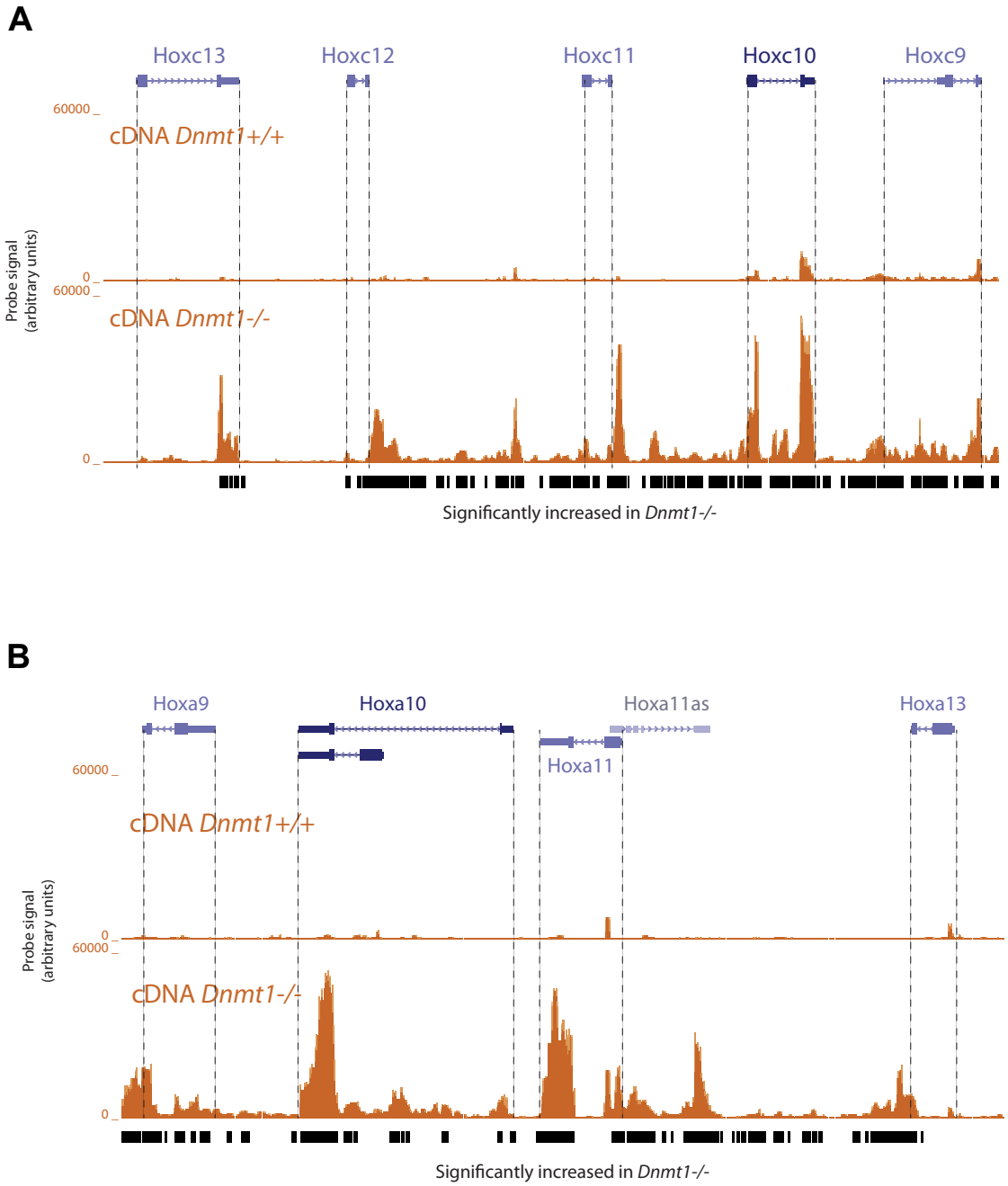


al., 2011; Bertani *et al.*, 2011). Also, a 'sweeping' mechanism has been suggested where the act of ncRNA transcription itself causes dissociation of PcG complexes (Hekimoglu and Ringrose, 2009). As DNA methylation has a well established role in repression of transcription, I hypothesised that de-repression of ncRNA may contribute to K27me3 loss in *Dnmt1*$^{-/-}$ MEFs. To address this idea, I analysed the cDNA custom tiled microarray used in chapter 3 in more detail. This microarray consists of densely tiled probes covering all four *Hox* clusters and so can be used to identify changes in intergenic transcription. A similar approach was successfully used in human cells by Rinn *et al.*, (2007). I was aware of certain limitations while performing this experiment. Firstly, cDNA was synthesised using oligo-dT priming, meaning that not all RNA would be represented. Multiple previously identified ncRNAs within *Hox* clusters have been shown to be spliced and poly-adenylated so I felt this to be only a minor limitation (Rinn *et al.*, 2007; Wang *et al.*, 2011; Bertani *et al.*, 2011). Secondly, the microarray and cDNA synthesis technique used here results in loss of strand information. I reasoned that any ncRNA observed could be investigated and mapped in more detail subsequently. Thirdly, as with any cDNA experiment, the levels of steady state RNA are measured as appose to transcription itself.

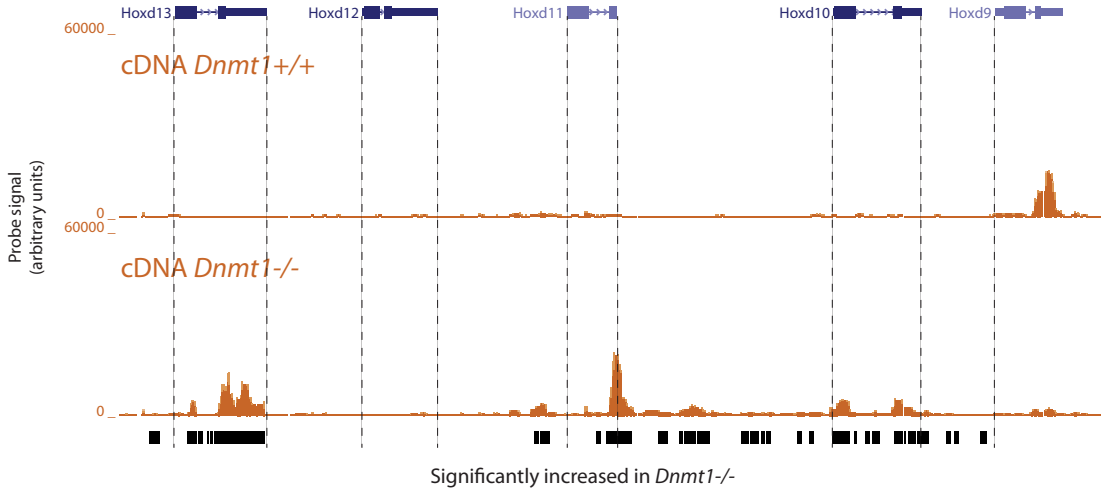
Further examination of cDNA custom tiled microarray results showed that extensive increases in intergenic transcription are observed within *Hox* clusters in *Dnmt1*$^{-/-}$ MEFs. I used the position of RefSeq genes to identify genic and intergenic regions. Probes with significantly increased signal in *Dnmt1*$^{-/-}$ MEFs were identified by finding groups of three consecutive probes with $p<0.05$ by two-tailed Student's t-test (see chapter 2 for details). Significantly increased signal was observed at the *HoxC*, *HoxA* and *HoxD* clusters in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.9A to C). An annotated non-coding antisense transcript from the *HoxA* cluster, *Hoxa11as* (*Hoxa11* antisense), also showed significant upregulation in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.9B). The *HoxD* cluster shows less widespread increase in intergenic cDNA signal than the *HoxA* and *HoxC* clusters, correlating with the less severe reductions in K27me3 at this cluster in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.9A to C and chapter 4). As mentioned in chapter 4, no probes in the *HoxB* cluster show increased signal in *Dnmt1*$^{-/-}$ MEFs. I calculated the proportion of *Hox* clusters that show significantly increased signal in *Dnmt1*$^{-/-}$ MEFs. About 20% and 40% of the probes in the *HoxA* and *HoxC* clusters respectively are significantly upregulated in *Dnmt1*$^{-/-}$ MEFs, compared to around 3% of all probes on this array (Fig. 5.9D). Similar proportions were observed using only intergenic probes (Fig. 5.9E). These results suggest that many intergenic ncRNAs from within *Hox* clusters are upregulated in *Dnmt1*$^{-/-}$ MEFs.

Figure 5.9 - Profiling cDNA using tiling arrays reveals increased intergenic transcription in *Hox* clusters in *Dnmt1*$^{-/-}$

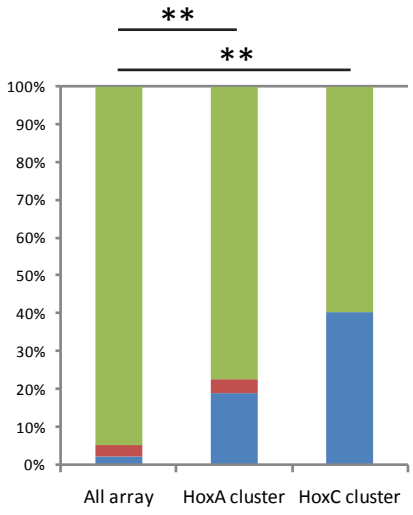
A. UCSC browser images showing cDNA array signals (custom tiled array) for a portion of the *HoxC* cluster in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. RefSeq genes are shown in blue. Dashed vertical lines indicate the start and stop positions of RefSeq genes. Black bars (bottom) indicate probes with significantly increased signals in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs (see Methods). B. As 'A' but showing data for a portion of the *HoxA* cluster. C. As 'A' but for a portion of the *HoxD* cluster. D. Proportion of probes with significantly different cDNA signals in *Dnmt1*$^{-/-}$ MEFs within *Hox* clusters. ** denotes p<0.0001 by chi-squared test between indicated samples. Note that the key (right) is for both 'D' and 'E'. E. As 'D' but for intergenic probes only.



C

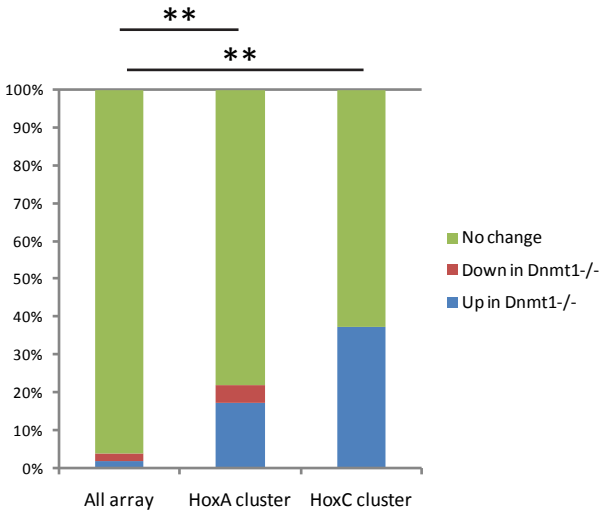


D



All probes

E



Intergenic probes

Although ncRNA transcription has been linked with gene activation and relief from PcG repression, how it does this remains unclear (Hekimoglu and Ringrose, 2009). Recently, the H3K36me3 histone mark has been suggested to be a potent PRC2 antagonist (Schmitges *et al.*, 2011; Yuan *et al.*, 2011). The H3K36me3 and K27me3 marks rarely exist on the same nucleosomes in human cells and, *in vitro*, activity of the PRC2 complex on nucleosomes is inhibited by pre-installed H3K36me3 (Schmitges *et al.*, 2011; Yuan *et al.*, 2011). As H3K36me3 is a mark associated with transcription units, specifically, elongating RNA polymerase II, it has been postulated that this antagonism prevents K27me3 spreading into actively transcribed regions. I hypothesised that intergenic transcription may contribute to PRC2 displacement/inhibition by increasing the level of intergenic H3K36me3 within *Hox* clusters. To begin to test this hypothesis I performed N-ChIP for the H3K36me3 mark in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. I confirmed that the N-ChIP was successful by assaying enrichment at control regions by N-ChIP-qPCR (Fig. 5.10A). The *Gapdh* promoter and exonic region would be expected to be negative and positive respectively in both cell types, and this pattern was observed by N-ChIP-qPCR (Fig. 5.10A). The exonic region of *Tex19.1* would be expected to be positive in *Dnmt1*$^{-/-}$ and negative in *Dnmt1*$^{+/+}$ MEFs, where the gene is expressed and repressed respectively. As N-ChIP for H3K36me3 appeared to have been successful, I labelled the DNA and hybridised to the custom tiled microarray. After normalisation, I checked the enrichment of all exonic and intergenic probes in *Dnmt1*$^{+/+}$ MEFs. As expected, exonic probes showed an overall higher signal while intergenic probes showed an overall lower signal than all probes (Fig. 5.10B, $p<0.0001$ by Mann-Whitney U test). Also as expected, exonic probes show a positive relationship between cDNA signal and K36me3, although considerable variation was observed suggesting that H3K36me3 and expression are not constitutively tightly correlated (Fig. 5.10C). Overall, these results suggest that N-ChIP-chip for H3K36me3 performed here can effectively be used to profile levels of this histone mark.

The signal for H3K36me3 N-ChIP-chip was compared between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs (Fig. 5.11A). Overall, a greater variation between the two cell types was observed for probes of the *HoxA* and *HoxC* clusters than for all probes, as assessed by Pearson's correlation co-efficients (Fig. 5.11A, R=0.851: All probes, R=0.571: *HoxC*, R=0.614: *HoxA*). In particular, many probes from the *HoxC* and *HoxA* clusters show increased signal in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs (Fig. 5.11A). In *Dnmt1*$^{+/+}$ MEFs, a clear partition in H3K36me3 signal, as observed for cDNA, was not apparent (see *HoxC* cluster in Fig. 5.11B for example). This further suggests that H3K36me3 and cDNA signal are not always tightly correlated. However, by examining the signal within these clusters it

Figure 5.10- Mapping H3K36me3 using custom tiled arrays

A. ChIP-qPCR for contol regions in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Error bars indicate standard error of the mean. B. Boxplot showing mean signals from the tiling array in *Dnmt1*$^{+/+}$ MEFs for exonic, intergenic and all probes. ** denotes p<0.00000001 by Mann-Whitney-U test between indicated groups. C. Scatter plot showing correlation between K36me3 and cDNA signal in *Dnmt1*$^{+/+}$ MEFs at exonic probes. Blue line is plotted by lowess. R-value was calculated by Pearson's correlation.

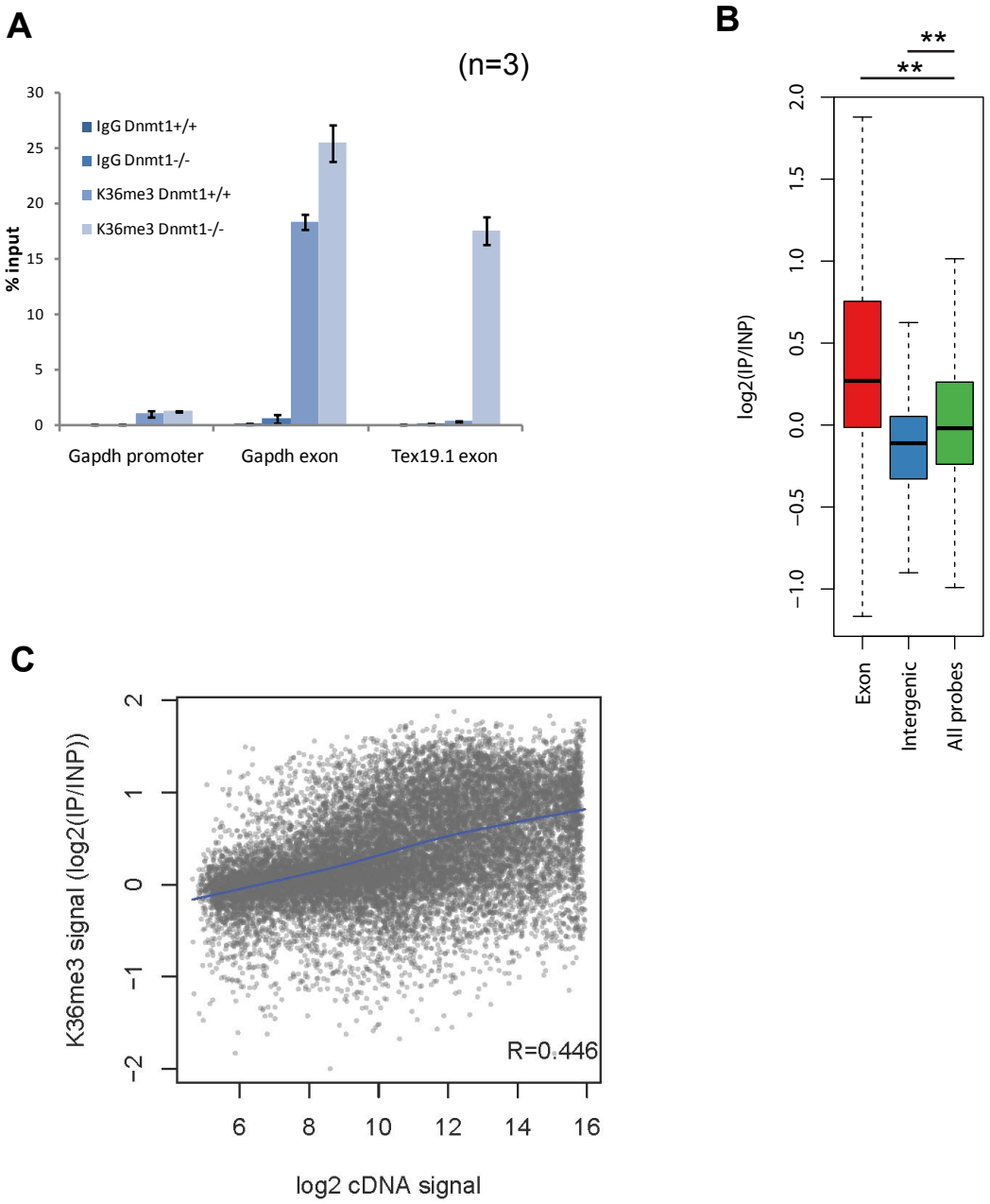
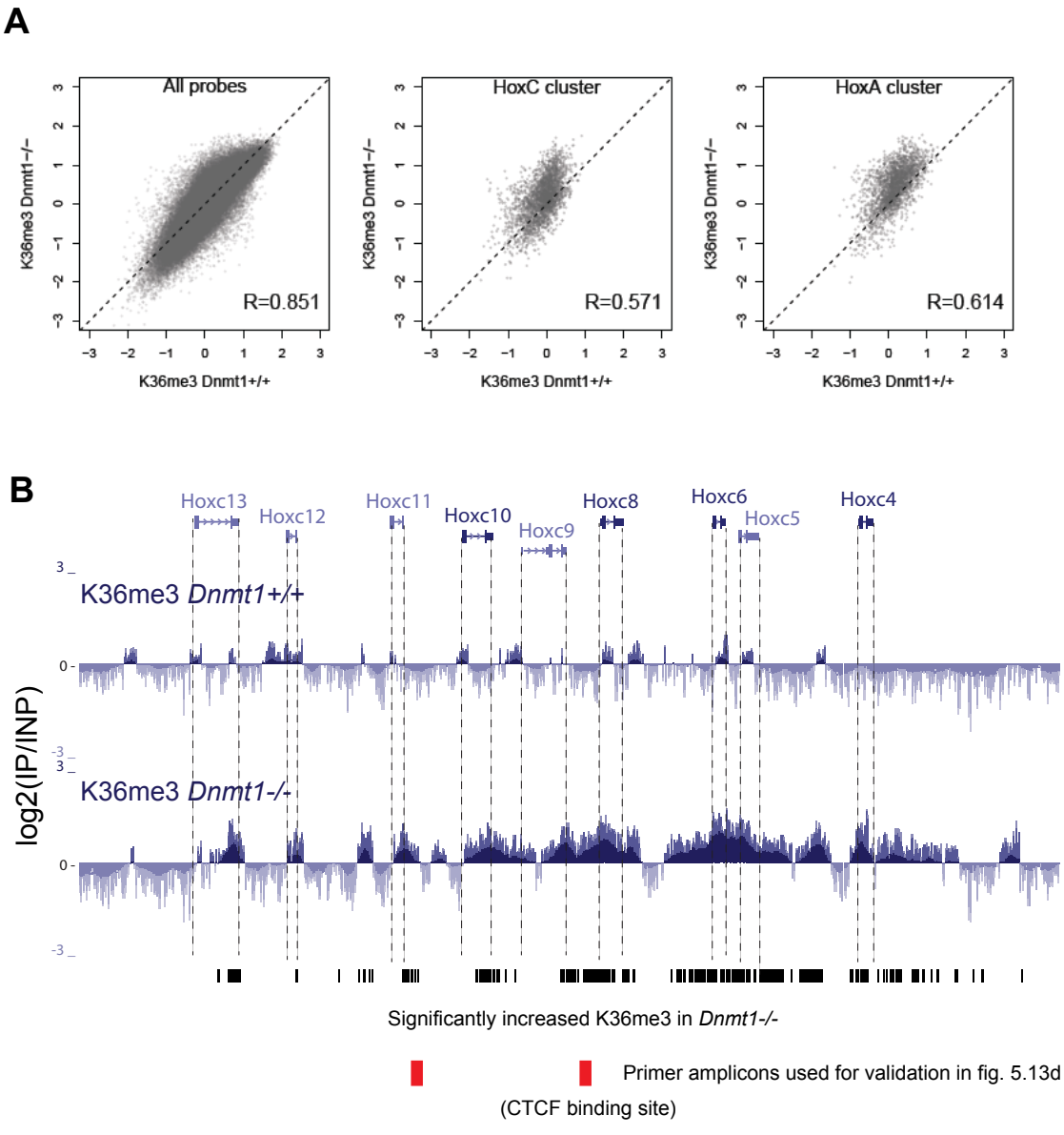
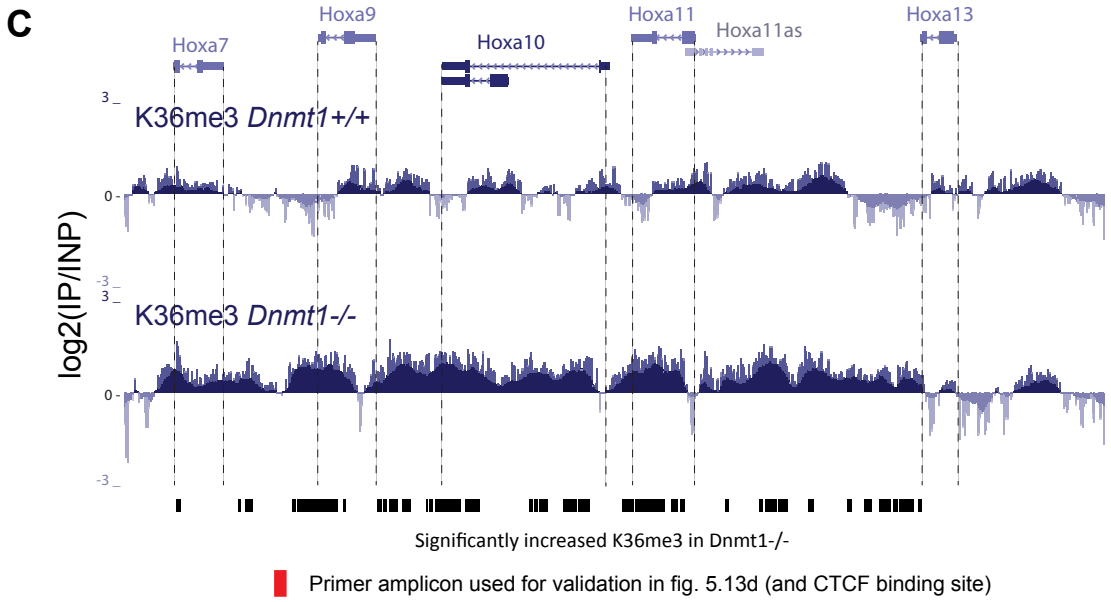
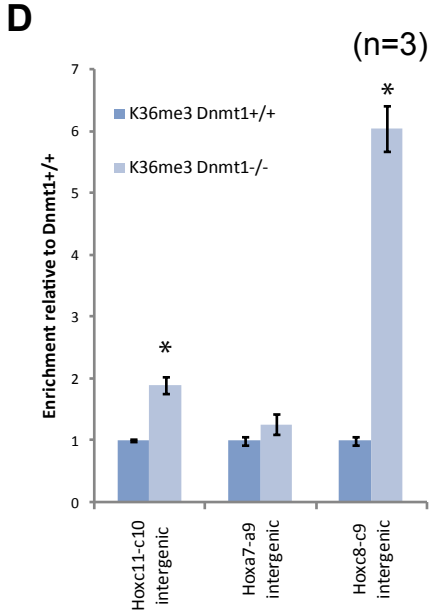


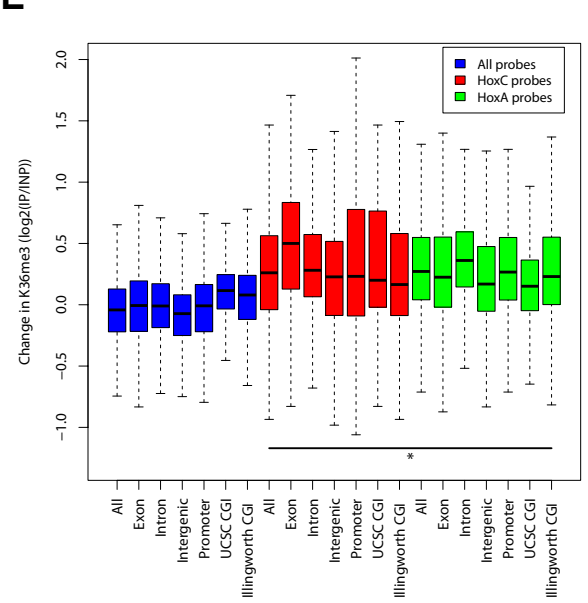
Figure 5.11 - H3K36me3 at *Hox* clusters increases in the absence of *Dnmt1*

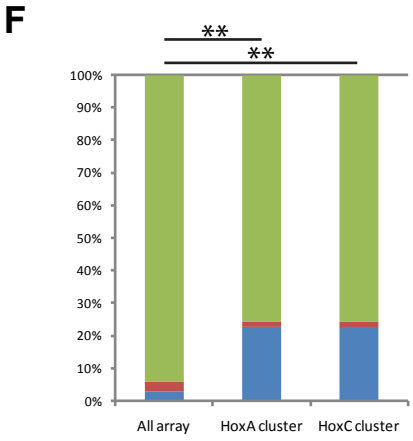
A. Scatter plots of K36me3 signal as measured by ChIP-chip on *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs at *Hox* clusters and all probes. Dashed lines are y=x. R-values are from Pearson's correlations. B. UCSC browser image showing K36me3 signal at the *HoxC* cluster in *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. Position of RefSeq genes are shown in blue. Dashed vertical lines indicate start and stop positions of RefSeq genes. Black bars (bottom) indicate probes that show significant increase in signal in *Dnmt1*$^{-/-}$ MEFs (see Methods). Amplicon of primers used in 'D' are shown as red blocks (bottom). C. As 'B' but showing a portion of the *HoxA* cluster. D. ChIP-qPCR validation of ChIP-chip data for two regions that show increased K36me3 in *Dnmt1*$^{-/-}$ MEFs and one region that shows no change. Error bars represent +/- standard error of the mean (n=3). Primer binding locations are shown in 'B' and 'C'. * denotes p<0.05 by Student's t-test. E. Boxplot showing difference in K36me3 between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs as measured by ChIP-chip (custom tiled array) at various annotations within *Hox* clusters. * indicates that all annotations tested within *Hox* clusters have significantly higher K36me3 signal than the equivalent annotation using all probes (p<0.005, Mann-Whitney-U test). F. Proportions of *Hox* cluster probes that show significantly different K36me3 signal between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. ** denotes a p<0.0001 by chi-squared test between indicated groups. G. As 'F' but using only intergenic probes.

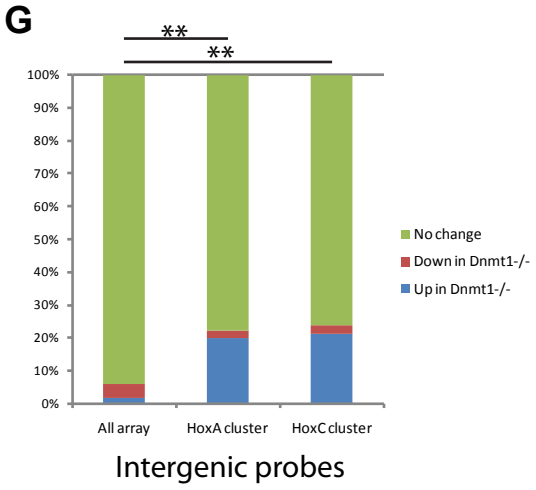



C
Hoxa7
Hoxa9
Hoxa10
Hoxa11
Hoxa11as
Hoxa13
K36me3 Dnmt1+/+
K36me3 Dnmt1-/-
log2(IP/INP)
Significantly increased K36me3 in Dnmt1-/-
Primer amplicon used for validation in fig. 5.13d (and CTCF binding site)


D
(n=3)
K36me3 Dnmt1+/+
K36me3 Dnmt1-/-
Enrichment relative to Dnmt1+/+
Hoxc11-c10 intergenic
Hoxa7-a9 intergenic
Hoxc8-c9 intergenic


E
All probes
HoxC probes
HoxA probes
Change in K36me3 (log2(IP/INP))
All
Exon
Intron
Intergenic
Promoter
UCSC CGI
Illingworth CGI


F
All array
HoxA cluster
HoxC cluster
All probes


G
No change
Down in Dnmt1-/-
Up in Dnmt1-/-
All array
HoxA cluster
HoxC cluster
Intergenic probes

is clear to see that, similar to cDNA, many intergenic regions show increased H3K36me3 signal in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.11B and C). I confirmed increased H3K36me3 in *Dnmt1*$^{-/-}$ MEFs at two of these loci by N-ChIP-qPCR (Fig. 5.11D, amplicon locations shown in Fig. 5.11B and C, $p<0.05$ by two-tailed Student's t-test). Of note, the tested regions included the CTCF binding sites examined in figure 5.7. By microarray, both of these regions show increased H3K36me3 in *Dnmt1*$^{-/-}$ MEFs, although this is observed in the flanking regions to the CBS5 CTCF binding site in the *HoxA* cluster, and so is not observed by N-ChIP-qPCR using these primers (Fig. 5.11B and C). Dividing *HoxA* and *HoxC* clusters into different annotations shows that increased H3K36me3 signal is observed at all of the analysed annotations in *Dnmt1*$^{-/-}$ MEFs, including intergenic and promoter regions (Fig. 5.11E, $p<0.005$ by Mann-Whitney U test compared to all probes). Greater than 20% of probes from *HoxA* and *HoxC* clusters have significantly increased H3K36me3 signal in *Dnmt1*$^{-/-}$ MEFs while less than 4% of all probes show this trend (Fig. 5.11F, $p<0.0001$ by chi-squared test). A similar proportion of intergenic probes show significantly increased signal, suggesting that, like cDNA, H3K36me3 is increased in many intergenic regions of *Hox* clusters in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.11G).

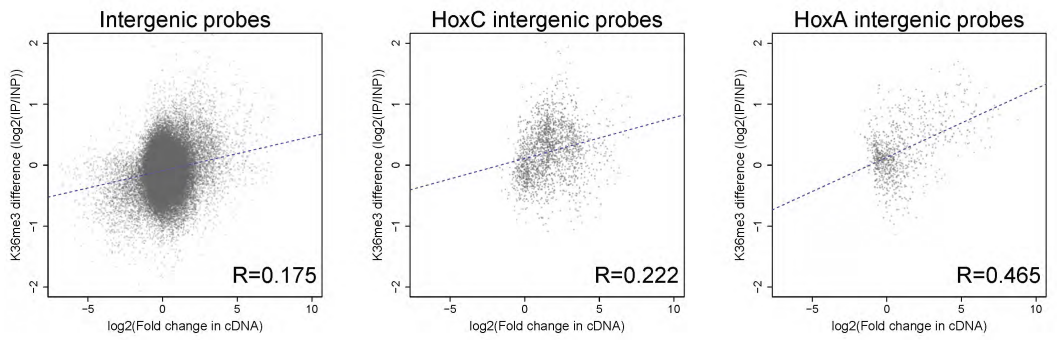
I next tested if changes in cDNA signal and H3K36me3 within *Hox* clusters are correlated. This would be expected if increases in transcription cause H3K36me3 increase. For all intergenic probes on the custom tiled array, a fairly weak correlation between cDNA difference and K36me3 difference (each between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs) is observed (Fig. 5.12A, R=0.175). Within the intergenic probes of the *HoxA* and *HoxC* clusters, a slightly better correlation is observed, but this correlation is still weak (Fig. 5.12A, R=0.222: *HoxC*, R=0.465: *HoxA*). This further suggests that these two signals are not constitutively associated. This could potentially be explained by the presence of transcripts that are not detected using this method, or additional factors that influence the H3K36me3 mark.

There are about 70 genes on the custom tiled array that are K27me3 targets in *Dnmt1*$^{+/+}$ MEFs (defined by having an average promoter log2(IP/INP) > 0.5) and many of these genes are suggested to lose promoter K27me3 in *Dnmt1*$^{-/-}$ MEFs (by promoter array, described in chapter 4). In order to test if increased intergenic transcription and H3K36me3 are associated with K27me3 loss, I plotted a heatmap of the ~70 K27me3 target genes (Fig. 5.12B). I scored each gene for whether they are associated with increased cDNA or H3K36me3 signal (three consecutive probes with $p<0.05$ by Student's t-test) in exonic, intronic and intergenic regions. Intergenic regions associated with each gene were defined as up to 3kb away and not within any other RefSeq transcript. The heatmap was ordered by the difference in K27me3 between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs. The threshold for K27me3

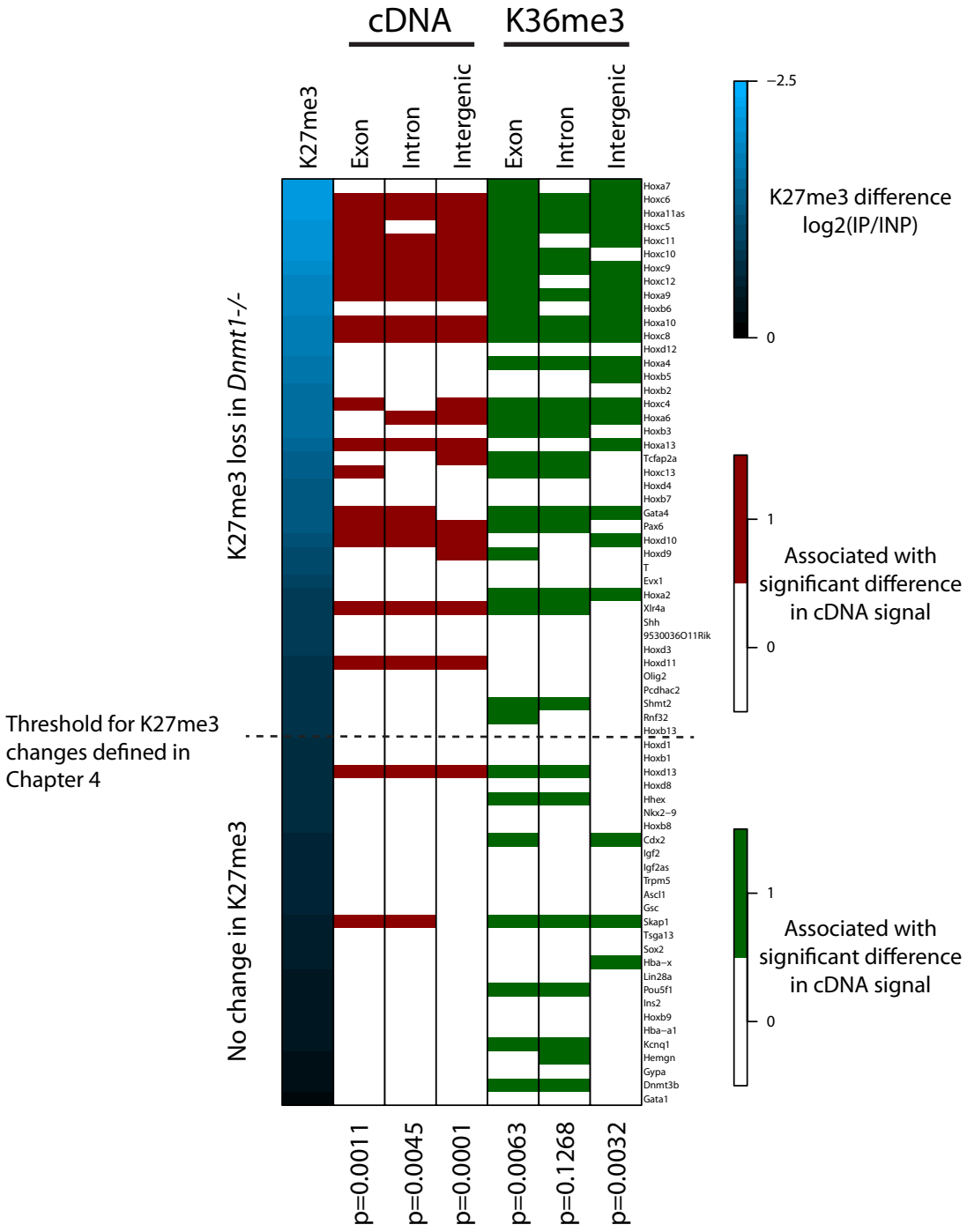
Figure 5.12 - Intergenic transcription and K36me3 are correlated and both are associated with K27me3 loss in *Dnmt1*$^{-/-}$

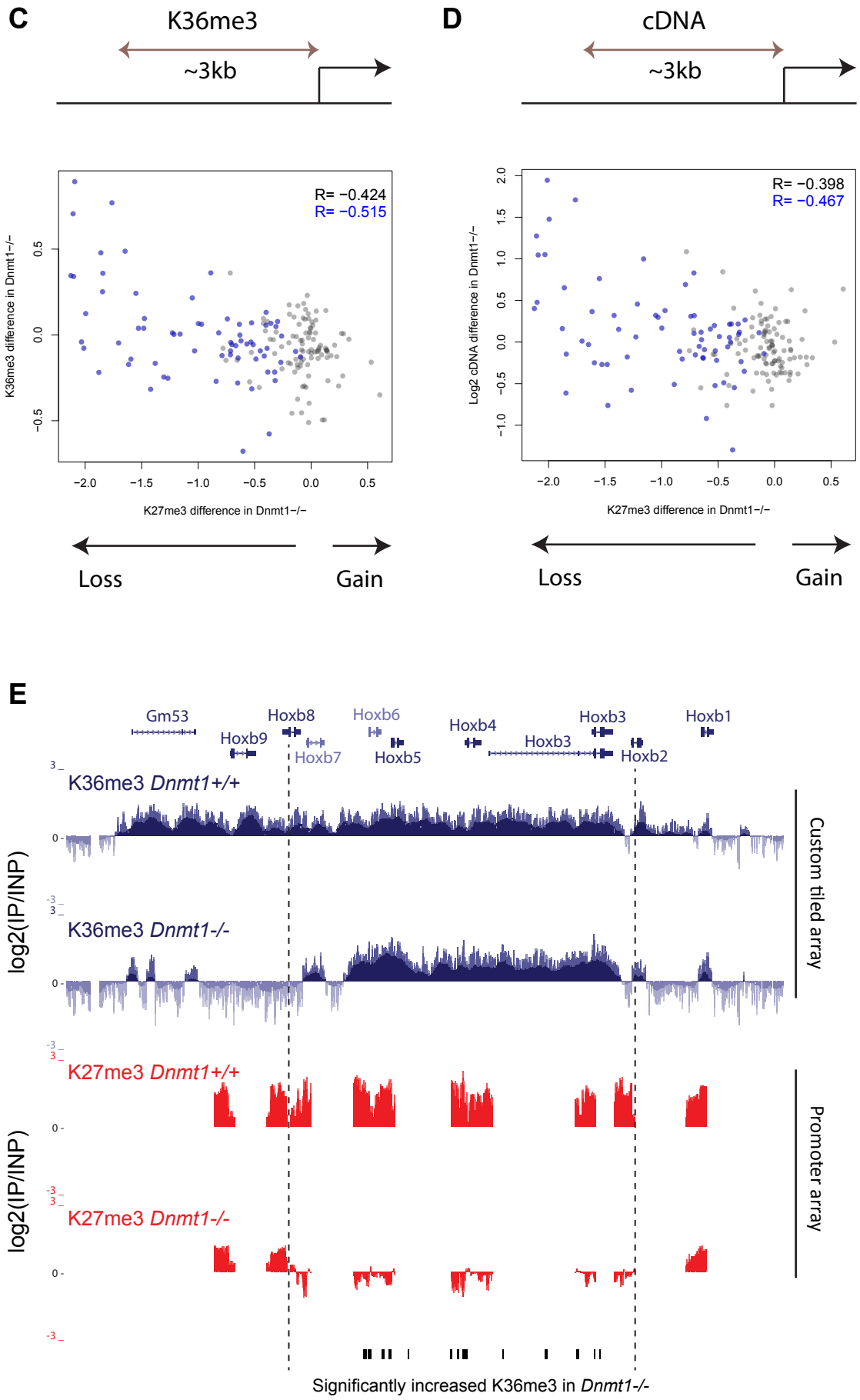
A. Scatter plots showing relationship between cDNA signal difference and K36me3 ChIP signal difference (between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$) for intergenic probes on tiling array. Dashed blue line was fitted by regression. R-value was calculated by Pearson's correlation. B. Heatmap showing association of K27me3, K36me3 and cDNA changes at all K27me3 targets on the custom tiled microarray . K27me3 panel shows difference in K27me3 at gene promoters as measured by N-ChIP-chip (promoter array). cDNA (red) and K36me3 (green) panels show if the gene is associated with probes from the custom tiled array which show significantly increased signal in *Dnmt1*$^{-/-}$. Intergenic was defined as <3kb from TSS and not within a RefSeq gene. The dashed line represents the threshold for K27me3 loss defined in chapter 4. The p-values (bottom) are from a Fisher's exact test between groups defined by the K27me3 loss threshold. The gene names are shown to the right of the heatmap. C. Scatter plot for K36me3 difference and K27me3 difference between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs for all genes covered by the custom tiled array. K36me3 difference was calculated by taking an average of the non-genic promoter-associated probes for each gene from the custom tiled array. K27me3 difference comes from the promoter array described in chapter 4. R-value is from a Pearson's correlation. Genes marked by K27me3 in *Dnmt1*$^{+/+}$, and the R-value for these points, are plotted in blue. D. As 'C' but for non-genic promoter cDNA signal. E. UCSC browser image showing N-ChIP-chip data for K27me3 (promoter array) and K36me3 (custom tiled array) at the *HoxB* cluster. The position of RefSeq genes are shown (blue). Black bars (bottom) indicate probes that are significantly increased in K36me3 signal in *Dnmt1*$^{-/-}$. Dashed vertical lines indicate region of cluster that shows decreased K27me3 in *Dnmt1*$^{-/-}$.

A



B


cDNA
K36me3
K27me3
Exon
Intron
Intergenic
Exon
Intron
Intergenic
K27me3 loss in Dnmt1-/-
No change in K27me3
Threshold for K27me3 changes defined in Chapter 4
Hoxa7
Hoxc6
Hoxa11as
Hoxc5
Hoxc11
Hoxc10
Hoxc9
Hoxc12
Hoxa9
Hoxb6
Hoxa10
Hoxc8
Hoxd12
Hoxa4
Hoxb5
Hoxb2
Hoxc4
Hoxa6
Hoxb3
Hoxa13
Tcfap2a
Hoxc13
Hoxd4
Hoxb7
Gata4
Pax6
Hoxd10
Hoxd9
T
Evx1
Hoxa2
Xlr4a
Shh
9530036O11Rik
Hoxd3
Hoxd11
Olig2
Pcdhac2
Shmt2
Rnf32
Hoxb13
Hoxd1
Hoxb1
Hoxd13
Hoxd8
Hhex
Nkx2−9
Hoxb8
Cdx2
Igf2
Igf2as
Trpm5
Ascl1
Gsc
Skap1
Tsga13
Sox2
Hba−x
Lin28a
Pou5f1
Ins2
Hoxb9
Hba−a1
Kcnq1
Hemgn
Gypa
Dnmt3b
Gata1
p=0.0011
p=0.0045
p=0.0001
p=0.0063
p=0.1268
p=0.0032
−2.5
0
K27me3 difference log2(IP/INP)
1
0
Associated with significant difference in cDNA signal
1
0
Associated with significant difference in cDNA signal


C
K36me3
~3kb
R= −0.424
R= −0.515
K36me3 difference in Dnmt1−/−
K27me3 difference in Dnmt1−/−
Loss
Gain
D
cDNA
~3kb
R= −0.398
R= −0.467
Log2 cDNA difference in Dnmt1−/−
K27me3 difference in Dnmt1−/−
Loss
Gain
E
Gm53
Hoxb9
Hoxb8
Hoxb7
Hoxb6
Hoxb5
Hoxb4
Hoxb3
Hoxb3
Hoxb2
Hoxb1
K36me3 Dnmt1+/+
K36me3 Dnmt1-/-
K27me3 Dnmt1+/+
K27me3 Dnmt1-/-
log2(IP/INP)
log2(IP/INP)
Custom tiled array
Promoter array
Significantly increased K36me3 in Dnmt1-/-

loss as defined in chapter 4 is shown on the heatmap as a horizontal dashed line. Genes above this line are defined as having lost K27me3 in *Dnmt1*$^{-/-}$ MEFs, allowing statistical testing of any association by categorical Fisher's exact test. Genes at the top of this heatmap, those which show the most K27me3 loss in *Dnmt1*$^{-/-}$ MEFs, are frequently associated with increased cDNA and H3K36me3 in exonic, intronic and intergenic regions (Fig. 5.12B). This observation is confirmed by Fisher's exact test for each annotation. The lowest p-value by Fisher's exact test, for both cDNA and H3K36me3, was observed for intergenic regions, suggesting that intergenic changes in these features correlate better with K27me3 loss than gene-associated changes, for these genes using this analysis (Fig. 5.12B).

I addressed this idea using a second analysis. For each gene on the custom tiled array, I calculated the mean probe signal change for cDNA and H3K36me3, between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ MEFs, within upstream promoter regions. Each upstream promoter region was defined as up to 3kb upstream of the gene transcription start site and not within another RefSeq transcript. This provides only non-coding, intergenic probes for each region. The difference in H3K36me3 and cDNA within this upstream region for each gene was calculated and plotted against the K27me3 change associated with each gene from promoter arrays. This allows the promoter-associated H3K36me3 and cDNA change to be correlated with promoter K27me3 changes for these genes. Increase of either H3K36me3 or cDNA in upstream promoter regions is correlated with K27me3 loss in *Dnmt1*$^{-/-}$ MEFs (Fig. 5.12C and D, R=0.424: K36me3, R=0.398: cDNA). This analysis contains all genes with sufficient coverage on the custom tiled microarray. Plotting only the ~70 genes that are K27me3 targets in *Dnmt1*$^{+/+}$ MEFs (in blue), also showed a correlation, with slightly better R values (Fig. 5.12C and D, R=0.515: K36me3, R=0.467: cDNA). These observations suggest that promoter-associated transcription and H3K36me3 are increased at many of the assayed genes in *Dnmt1*$^{-/-}$ MEFs. Increased promoter-associated cDNA and H3K36me3 correlates with loss of the PRC2-mediated K27me3 modification from these gene promoters. As H3K36me3 is a suggested PRC2 antagonist, these observations are consistent with a model where intergenic transcription contributes to loss of PRC2-mediated gene repression in *Dnmt1*$^{-/-}$ MEFs, although of course does not prove it. At this point it is equally likely that the observed increases in cDNA and H3K36me3 are a consequence of loss of PRC2-mediated repression. Further, more functional, studies are required to differentiate between the two possibilities. It is interesting to note that decreased K27me3 is observed at the *Hoxb2-b7* region of the *HoxB* cluster despite no observed increase in cDNA in this region (Fig. 5.12E, cDNA was shown in chapter 4). This region does, however, show increased H3K36me3 in *Dnmt1*$^{-/-}$ MEFs, despite decreased H3K36me3 in flanking regions (Fig. 5.12E). This could

indicate that transcription does not cause increased K36me3 and K27me3 loss, at least at this locus, or that transcripts that are not detected (for example those that are not poly-adenylated), are upregulated. Further study of this locus could help to differentiate between the two possibilities.

5.3. Discussion

In this chapter I have begun to investigate the mechanisms by which DNA methylation contributes to *Hox* gene regulation. By doing this I also aimed to better understand the role that DNA methylation plays in targeting the K27me3 histone modification to chromatin, a function that was suggested from my work presented in chapters 3 and 4. A discussion of the outcome of these experiments is presented under the following headings.

5.3.1. The DNA methylation pattern at Hox genes and its role in gene regulation

I first set out to characterise the DNA methylation pattern within *Hox* clusters in the type of cells that were used as the primary model during this work, mouse embryonic fibroblasts. I used more than one method to investigate DNA methylation patterns, as many of the existing high throughput techniques have certain limitations (Laird *et al.*, 2010). I used MeDIP-chip to profile DNA methylation at *Hox* clusters in *Dnmt1*$^{+/+}$ MEFs. Using this technique I established that promoter regions of *Hox* genes contain an 'intermediate' level of enrichment in *Dnmt1*$^{+/+}$ MEFs. In fact, a peak in enrichment was observed at nearly all CpG islands using this technique in DNA from these cells. The MeDIP technique is subject to significant biases, for example in CpG density, which can lead to misleading signals (Down *et al.*, 2008; Pelizzola *et al.*, 2008). For example, even low level DNA methylation in a very CpG-rich region can produce an enrichment in signal relative to the highly methylated, but CpG-poor, flanking regions. Techniques have been developed to try to account for these biases in MeDIP (Down *et al.*, 2008; Pelizzola *et al.*, 2008). For future MeDIP-chip experiments, one of these correction methods could be used to make the output more intuitive. For example, MeDIP on a genome that has been *in vitro* methylated to completion using the SssI CpG methyltransferase can be performed. When performed in parallel with samples, this can be

used to model the relationship between enrichment and CpG density, which in turn can be used to correct data from parallel MeDIP-chip experiments (Pelizzola *et al.*, 2008).

In this chapter, I performed bisulfite sequencing as an independent method to verify DNA methylation at certain loci. This combination of a relatively large scale method, MeDIP-chip, with the smaller scale, but more quantitative, bisulfite sequencing, was designed to provide reliable DNA methylation data within *Hox* clusters. Bisulfite sequencing of *Hox* promoters confirmed that they contain an 'intermediate' level of DNA methylation. This in itself could be considered somewhat surprising as the frequency distribution of CpG methylation in mammals is reportedly bimodal, with the majority of CpGs being either frequently or infrequently DNA methylated (Meissner *et al.*, 2008). However, a similar observation has been reported for *Hox* gene promoter CpGs in MEFs, and also in some primary tissues (Xi *et al.*, 2007; Tao *et al.*, 2010). I also wanted to establish DNA methylation patterns in intergenic regions, particularly at the many orphan CpG islands within *Hox* clusters that have been reported to show tissue-specific methylation patterns (Illingworth *et al.*, 2008). These regions also showed an intermediate DNA methylation in *Dnmt1*$^{+/+}$ MEFs.

The intermediate patterns of DNA methylation that I observed prompted me to investigate the DNA methylation status of some of the same regions in primary MEFs. I felt that this was important given that DNA methylation patterns have been suggested to change in cell culture, and K27me3 marked genes are particularly susceptible to hypermethylation in a variety of situations (Meissner *et al.*, 2008; Rakyan *et al.*, 2010). DNA methylation was observed to be much reduced in primary MEFs at nearly all regions analysed. This result means that currently, I cannot draw any conclusions about the DNA methylation pattern of *Hox* clusters from *Dnmt1*$^{+/+}$ MEFs. Further investigation is required to determine if the DNA methylation observed in *Dnmt1*$^{+/+}$ MEFs indeed represents hypermethylation accrued in culture. If this turns out to be the case then previously published work on DNA methylation in *Hox* clusters using cells in culture will also have to be reconsidered. Further study using primary tissues would be useful here. A further complication to this analysis is the observation of some regions within *Hox* clusters that do not show the expected hypomethylation in *Dnmt1*$^{-/-}$ MEFs, and the couple of regions that even have increased methylation in these cells. As Dnmt1 levels are thought to be severely decreased in these cells, presumably, this effect is mediated by increased targeting of *de novo* methyltransferases to these regions (Li *et al.*, 1992; Hackett, 2010, PhD thesis). As the processes required to target these proteins in normal cells is not well understood, the cause behind this effect remains obscure. One could speculate that increased H3K36me3 and

ncRNA transcription, as both have been linked to targeting DNMT3 proteins, could mediate this effect (Schmitz *et al.*, 2010; Dhayalan *et al.*, 2010). However, as H3K36me3 and cDNA increases were observed to occur at many regions in *Hox* clusters, it is unclear why DNA methylation should be increased at only specific regions if this were the cause. Another possibility that cannot be formally excluded is the presence of 5hmC which cannot be differentiated from 5mC using bisulfite sequencing (Nestor *et al.*, 2010).

Currently, a detailed knowledge of the normal DNA methylation pattern at *Hox* clusters represents a significant gap in our understanding of the mechanisms at play here. Therefore, immediate future work should focus on this aspect in order to better understand the putative role of DNA methylation in K27me3 targeting. MeDIP-chip could be performed using DNA from primary MEFs and/or tissue to begin to address this issue.

5.3.2. Involvement of CTCF and chromatin loops?

A number of observations led up to the investigation of CTCF protein binding to putative boundary elements within *Hox* clusters. For example, an apparent spreading of the K4me3 histone mark and cDNA signal into the normally repressed *Hox9-13* domain was observed at three of the four *Hox* clusters in *Dnmt1*$^{-/-}$ MEFs. CTCF was observed to be bound to an element between the *Hoxc8* and *Hoxc9*, and *Hoxa7* and *Hoxa9* (there is no *Hox8* gene in the *HoxA* cluster), genes in *Dnmt1*$^{+/+}$ MEFs when the two halves of these clusters appear to be partitioned. The boundary element activity of this region has been tested by Kim *et al.* (2011) where it was shown to insulate a reporter gene from a silencing signal. Also, proper K27me3 and gene expression patterns at the *HOXA* locus reportedly depend on CTCF in human foetal lung fibroblasts (Kim *et al.*, 2011). Chromatin conformation capture (3C) assay suggested that this element interacts with a downstream region of the *HoxA* cluster, implying formation of an intra-chromosomal loop that may be important for the cluster partitioning in differentiated cells (Kim *et al.*, 2011)(Fig. 5.14). Interestingly, reduced CTCF association with this region was detected in *Dnmt1*$^{-/-}$ MEFs, where the cluster partition appears to be lost (Fig. 5.14). It would be interesting to examine this observation in more detail. Firstly, the X-ChIP assay would need to be repeated as only one experiment has been performed so far. Also, the global levels of CTCF would need to be checked by western blot, although the equal binding to control regions observed by X-ChIP suggests that global levels of this protein are not affected in *Dnmt1*$^{-/-}$ MEFs. No difference was observed in locus chromatin compaction by 2D FISH. It would be interesting to investigate if the CTCF binding site does indeed form an intra-chromosomal loop in MEFs by 3C assay, and what

happens to this interaction in *Dnmt1*$^{-/-}$ MEFs (model shown in Fig. 5.14). Importantly, functional experiments would be required to test if ablation of CTCF binding to this site recapitulates any of the observed changes at *Hox* clusters. This could be performed by depletion of CTCF or by expression of a dominant negative CTCF protein, as performed by Zhang *et al.* (2011). However, CTCF depletion reportedly leads to apoptosis, at least in the developing limb bud (Soshnikova *et al.*, 2011). Alternatively, to avoid global effects mediated by these methods, single CTCF binding sites may be disrupted by complementary locked nucleic acids, a novel approach reported by Ling *et al.* (2011).

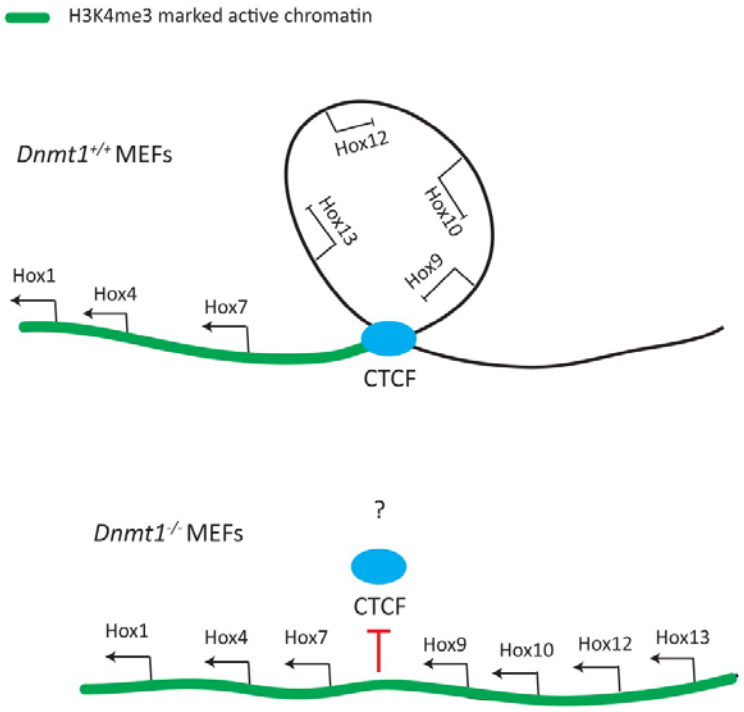


Figure 5.14 - CTCF-mediated partitioning of *Hox* clusters. A model for how a CTCF-mediated intra-chromosomal loop partitions a *Hox* cluster and prevents spreading of 'active' chromatin, represented by K4me3. Looping proposed by Kim *et al.* (2011). In *Dnmt1*$^{-/-}$ MEFs, binding of CTCF and the loop structure may be lost allowing spreading of active chromatin marks and activation of the normally repressed *Hox9-13* genes.

How would DNA methylation loss lead to loss of CTCF binding? This question would have to be addressed next if altered CTCF binding is a contributor to the observed effect of DNA hypomethylation on *Hox* cluster regulation. Although CpG methylation is thought to directly influence CTCF binding to certain cognate sites, it is suggested to have a negative effect. How DNA methylation would be required for CTCF binding is therefore unknown. Intriguingly, transcription of a ncRNA has been suggested to cause CTCF displacement in the activation of an LPS-induced gene (Lefevre *et al.*, 2008). In this respect, I observed an increase in cDNA and K36me3 signal in *Dnmt1*$^{-/-}$ MEFs, by custom tiled microarray, at the *Hoxc8-c9* CTCF binding site. However, a similar increase was not observed at the paralogous elements in the *HoxA* and *HoxD* clusters, although increased H3K36me3 was observed in regions flanking this element at the *HoxA* CBS5 site. Further investigation into this possibility is required.

5.3.3. Involvement of ncRNA?

In this chapter, I correlate increased expression of intergenic ncRNA with loss of K27me3 in *Dnmt1*$^{-/-}$ MEFs. Transcription through a PRC2 binding site has been linked with displacement of the complex and stable activation of nearby genes (Cavalli and Paro, 1998; Cavalli and Paro, 1999; Schmitt *et al.,* 2005; Sanchez-Elsner *et al.,* 2006). How ncRNA transcription influences chromatin regulation has been a mystery in recent years but mechanisms for ncRNA-modulation of chromatin modifications are beginning to emerge. A handful of studies have linked ncRNAs from mammalian *Hox* clusters to gene activation *in cis* (Rinn *et al.,* 2007; Dinger *et al.,* 2008; Wang *et al.,* 2011; Bertani *et al.,* 2011). Some of these ncRNAs are suggested to bind and recruit TrxG proteins, like Wdr5 and Mll1 (Dinger *et al.,* 2008; Wang *et al.,* 2011; Bertani *et al.,* 2011). In fly, induction of transcription across a PRE using an exogenous promoter relieves PcG-mediated silencing (Schmitt *et al.,* 2005). A 'sweeping' mechanism has been proposed to explain this where the act of transcription itself results in dissociation of PcG complexes as chromatin is disassembled (Hekimoglu and Ringrose, 2009). Although the mechanism of action is still being established, the large number of studies that now link ncRNA transcription to gene activation *in cis* suggests a bona fide role for this type of gene regulation strategy. Motivated by these studies, I hypothesised that ncRNA transcription could be a factor influencing K27me3 profiles at *Hox* genes in *Dnmt1*$^{-/-}$ MEFs. This could occur in two ways; 1) DNA methylation may directly repress some of these transcripts and, hence, in this way be required for a correct pattern of chromatin marks and regulation of *Hox* clusters, or, 2) ncRNA may be increased in hypomethylated cells indirectly from the DNA methylation mark and contribute to PRC2 displacement at certain loci. I observed striking and abundant increases in intergenic transcription in *Dnmt1*$^{-/-}$ MEFs which correlate with K27me3 loss for genes represented on the custom tiled array. I observed upregulation of cDNA signal at the annotated ncRNA *Hoxa11as*, but did not find a similar increase in regions around the ncRNAs reported to interact with TrxG proteins, such as *Mistral* (Bertani *et al.,* 2011) and *HOTTIP* (Wang *et al.,* 2011) despite these RNAs being reportedly poly-adenylated.

The transcription body associated histone mark H3K36me3 has recently been shown to inhibit the activity of the PRC2 complex *in vitro* (Schmitges *et al.,* 2011). I reasoned that increased intergenic transcription may inhibit PRC2, at least partially, through this histone modification. Consistent with this, H3K36me3 was observed to be increased in intergenic regions and also correlates with K27me3 loss (model shown in Fig. 5.15). However,

whether this data is consistent with a major role for intergenic transcription and H3K36me3 in K27me3 loss in *Dnmt1*$^{-/-}$ MEFs is unclear. For example, a considerable amount of H3K36me3 was observed at certain regions in *Dnmt1*$^{+/+}$ MEFs that are also marked by K27me3, such as the *HoxB* cluster and parts of the *HoxA* cluster. This suggests that PRC2 is able to establish K27me3 at these regions despite marking by H3K36me3. Also, the *HoxB* cluster shows loss of K27me3 from a defined region yet shows no evidence for increased intergenic transcription, in fact most of the cluster appears to be less expressed in *Dnmt1*$^{-/-}$ MEFs. This region does however show increased H3K36me3, despite flanking regions showing loss of this modification in *Dnmt1*$^{-/-}$ MEFs. Therefore, it remains unclear if this data is consistent with a universal role for intergenic transcription and H3K36me3 in K27me3 loss. It is possible that certain regions are important for PRC2 recruitment which would then allow spreading of this complex and its K27me3 mark, as has been suggested in flies for example at PRE elements (Schwartz *et al.*, 2010). This would mean that transcription at these specific regions would be important, rather than the clusters in general.

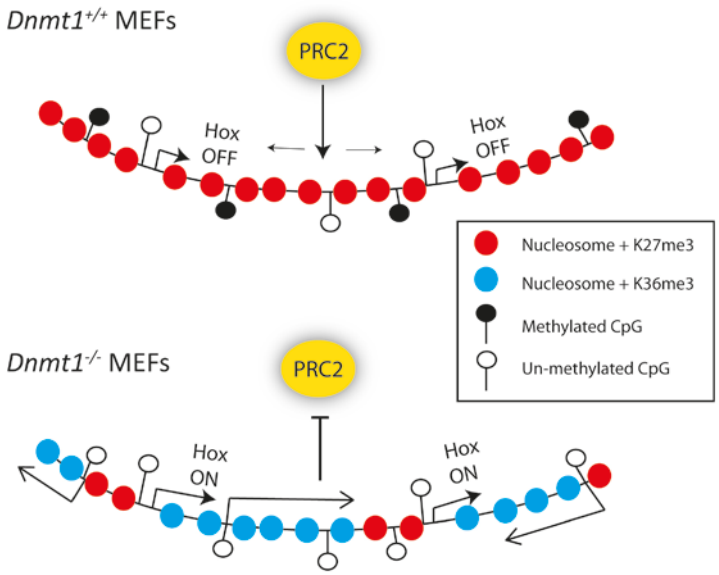


Fig. 5.15 - A model for how increased ncRNA transcription at *Hox* clusters could contribute to K27me3 loss in *Dnmt1*$^{-/-}$ MEFs. In *Dnmt1*$^{+/+}$ MEFs, intergenic DNA methylation prevents expression of intergenic transcripts and the region is bound by the PRC2 complex and marked by K27me3. Increased ncRNA transcription occurs in *Dnmt1*$^{-/-}$ MEFs and may directly interfere with PRC2 binding or inhibit PRC2 activity through increasing the transcription body associated H3K36me3 histone mark.

The limitation with this part of the project currently is that it is all correlative and no functional data has been generated. It is equally feasible that intergenic transcription and H3K36me3 increases are a consequence of loss of K27me3. These two possibilities will be difficult to separate as they would be expected to be intimately linked. Future experiments in this direction could involve the use of a transcription inhibitor on cells in culture. The

experiment would involve inhibition of transcription in *Dnmt1*$^{-/-}$ MEFs and assays for K27me3 level. This experiment is far from perfect as the time required for re-establishment of K27me3 marks is unknown and cells would not be expected to survive long in culture in the presence of a transcriptional inhibitor. Perhaps a better approach would be depletion of H3K36 methylases in *Dnmt1*$^{-/-}$ MEFs, although this would only test the role of this histone mark and not transcription *per se*. There are four methyltransferases in mammals that have been suggested to mediate H3K36 methylation (Table 5.2). Setd2, Nsd1 and Smyd2 are accepted as H3K36 methyltransferases but the *in vivo* substrate of Ash1l, the homolog of fly TrxG protein Ash1, is controversial (Yuan *et al.*, 2011). If H3K36me3 is required for normal *Hox* expression patterns then loss of the enzyme responsible would be expected to cause a developmental phenotype in mouse. Knockout of *Setd2* or *Nsd1* results in embryonic lethality and mouse embryos null for *Nsd1* fail to establish expression of *Hoxa1* and *Hoxb1* (Hu *et al.*, 2010; Rayasam *et al.*, 2003). *Smyd2* is dispensable for embryonic development so is unlikely to play a major role in H3K36me3 at *Hox* clusters (Diehl *et al.*, 2010). *Ash1l* has not been mutated in mouse as yet but interestingly, its homolog in fly has been linked to preventing the K27me3 mark (Papp and Mueller, 2006; Schwartz *et al.*, 2010). All of these genes are deemed to be significantly present in *Dnmt1*$^{+/+}$ MEFs by expression microarray (Table 5.2). Therefore, it may be interesting to deplete each of the Setd2, Nsd1 and Ash1l proteins in *Dnmt1*$^{-/-}$ MEFs and examine the effect on K27me3.

K36 methylase	**Knockout phenotype (mouse)**	**Expressed in *Dnmt1*$^{+/+}$ MEFs by expression microarray?**	**Refs**
Setd2	Lethal at ~10 dpc	Y	Hu *et al.*, 2010
Nsd1	Post-implantation lethal	Y (low)	Rayasam *et al.*, 2003
Smyd2	Dispensable; subtle heart and lung phenotypes	Y	Diehl *et al.*, 2010
(Ash1l)?	Not analysed	Y (low)	Yuan *et al.*, 2011

Table 5.2 - Proposed mammalian H3K36 methylases. The *in vivo* substrate of Ash1l is debated.

Whether the ncRNAs involved are directly repressed by DNA methylation remains unknown. To investigate this possibility, mapping of these transcripts, together with a better understanding of DNA methylation patterns at intergenic regions in *Hox* clusters, would be a

good starting point. More functional assays, such as region cloning and promoter assays coupled with *in vitro* CpG methylation, may prove informative.

Currently, I favour the idea that ncRNA transcription may contribute to loss of polycomb repression at certain loci, such as the *HoxA* and *HoxC* loci, and may explain why these regions are so drastically affected in *Dnmt1*$^{-/-}$ MEFs, but does not represent the primary cause of K27me3 loss. A general requirement of normal PRC2 targeting may be an intact DNA methylation pattern, leading to a redistribution of K27me3 at many loci in DNA hypomethylated cells. K27me3 loss may be exacerbated at certain loci, such as *HoxA* and *HoxC* clusters, because they are more sensitive to this effect. Increased intergenic transcription could contribute to this increased sensitivity, rather than represent the primary cause of K27me3 loss. Certain non-coding transcripts may also bind and recruit TrxG proteins to *Hox* genes, promoting their expression.

Chapter 6: General discussion

6.1. DNA methylation: A regulator of PRC2 targeting?

My aim for this project was to investigate novel functions of DNA methylation in genome regulation. I began by studying gene expression changes in a somatic cell line system that exhibits severe DNA hypomethylation. This led me to investigate the role of DNA methylation in the regulation of genes of the *Hox* clusters, genes encoding transcription factors with important roles in embryonic development. As *Hox* genes are major targets of the PcG repressor proteins, I became interested in the effect of DNA hypomethylation on the genomic distribution of the K27me3 histone mark, a surrogate for repression by the PRC2 complex. Finally, I returned to the *Hox* clusters to begin to study the mechanisms behind their mis-expression and the redistribution of K27me3 in DNA hypomethylated cells. The main findings of this work are as follows:

- Many PRC2 target genes are de-repressed in severely DNA hypomethylated *Dnmt1*$^{-/-}$ mouse embryonic fibroblasts (MEFs) and MEFs treated with hypomethylating agent 5-aza-dC.
- The organised regulation of *Hox* cluster genes is affected in *Dnmt1*$^{-/-}$ MEFs, co-incident with a striking loss of the PRC2-associated K27me3 histone modification and dissociation of the PRC2 component Ezh2.
- Posterior-expressed *Hox* gene regulation is also disrupted by treatment of anterior embryonic cells with demethylating agent 5-aza-dC and in *Dnmt1* mutant early mouse embryos.
- The distribution of the PRC2-associated K27me3 mark is severely altered in *Dnmt1*$^{-/-}$ MEFs, with loss of the mark from many gene promoters that are normally PRC2 targets, yet increase in the global levels of the mark. Preliminary evidence suggests that K27me3 may be re-targeted in *Dnmt1*$^{-/-}$ MEFs to some regions that normally have high levels of DNA methylation.
- K27me3 loss at *Hox* genes in *Dnmt1*$^{-/-}$ MEFs correlates with an apparent increase in intergenic transcription, as inferred by increased intergenic cDNA and transcription-unit-associated H3K36me3 histone modification.

The most interesting aspect of this work, in my opinion, is the link between DNA methylation and K27me3 targeting. At gene promoter regions I observed striking loss of the K27me3 mark in *Dnmt1*$^{-/-}$ MEFs at a large number of genes that are normally highly marked by this modification. By integrating promoter DNA methylation data, I suggest that the

majority of these genes are associated with a CpG island at their promoter region that contains low level DNA methylation in MEFs. Also, I observe a small number of promoter regions that show increased K27me3 in *Dnmt1*$^{-/-}$ MEFs and present preliminary evidence that these promoters are normally associated with DNA methylation. Despite a striking loss of K27me3 observed at many of the normally most highly marked promoters, globally an increase in the K27me3 modification was observed in *Dnmt1*$^{-/-}$ MEFs by western blot and immunofluorescence (IF). This poses an interesting question; where is the K27me3 modification in *Dnmt1*$^{-/-}$ MEFs? Addressing this question will be crucial to understanding the reason for K27me3 changes in these cells (discussed in the future work section). The observed increase in global K27me3 suggests that hypomethylation leads to increased levels of this mark at regions outside of gene promoters. This could potentially involve intergenic and gene body regions as both are thought to be highly DNA methylated in wildtype cells. I propose a working model for the mechanism that drives the observed K27me3 changes in DNA hypomethylated cells, based on redistribution of K27me3 through altered PRC2 binding throughout the genome (Fig. 6.1).

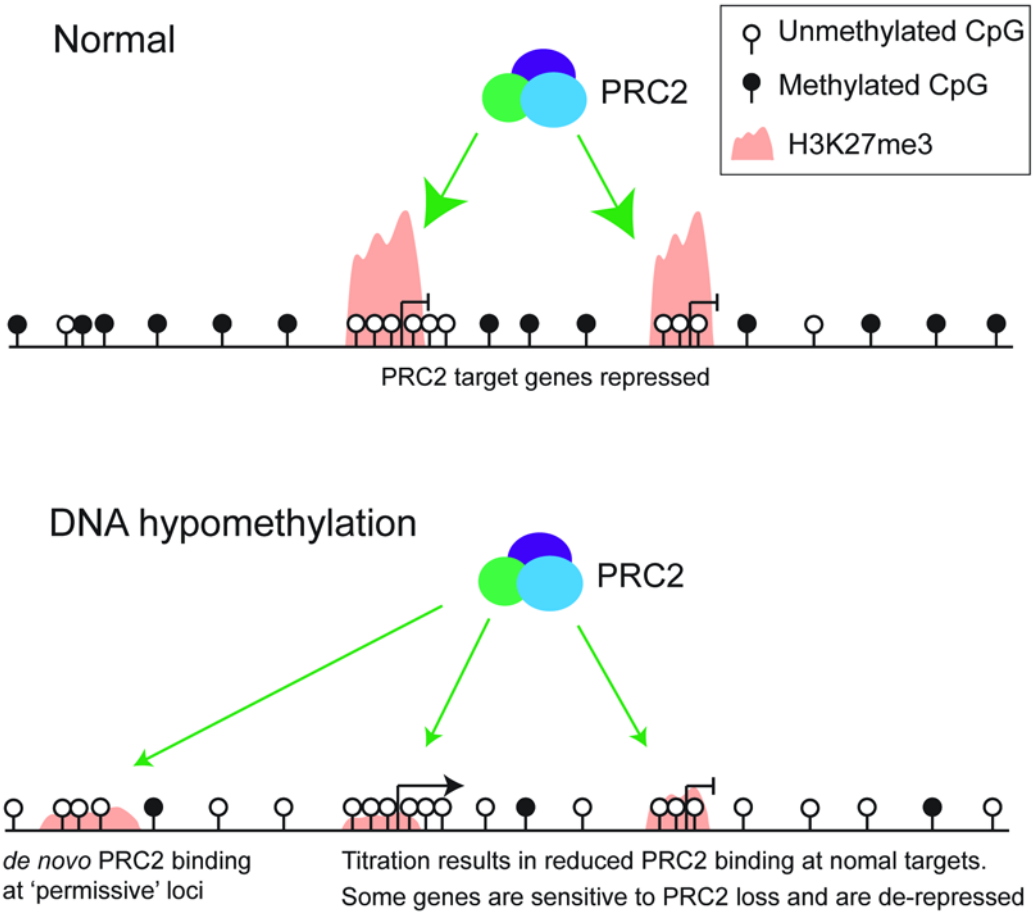


Fig. 6.1 - A working model for H3K27me3 redistribution in DNA hypomethylated cells. In normal cells PRC2 binding is inhibited by DNA methylation and is targeted by a process involving unmethylated CpG islands (based on data from Mendenhall *et al.*, 2010; Wu *et al.,* 2010; Bartke *et al.,* 2010; Lynch *et al.,* 2011). When DNA methylation is severely reduced, as in *Dnmt1*$^{-/-}$ MEFs, the K27me3 mark is lost from many polycomb-target gene promoters while global levels of this mark are not decreased. This implies that K27me3 is increased at regions outside of gene promoters. I propose that the depletion from normal targets is in part caused by titration of the PRC2 complex due to increased binding elsewhere in the genome. In this model, derepression of some PRC2-targets occurs as a consequence of loss of H3K27me3 and PRC2 binding.

Cytosine in CpG dinucleotides of the bulk genome are thought to be highly methylated in mammals, with stretches of low level DNA methylation thought to only occur at CpGs within CpG islands (Cooper *et al.*, 1983; Bird *et al.*, 1985; Meissner *et al.*, 2008; Lister *et al.*, 2009). Recent evidence suggests that PRC2 binding to chromatin is inhibited by DNA methylation, suggesting that the role of DNA methylation in K27me3 targeting may be a direct one (Wu *et al.*, 2010; Bartke *et al.*, 2010). Therefore, the 'global' DNA methylation pattern in mammals may help to restrict K27me3 patterns. Hypomethylation of bulk DNA in *Dnmt1*$^{-/-}$ MEFs may allow PRC2 binding at regions where it was previously inhibited. Only a small proportion of the genome is made up of CpG islands so it is possible that a small increase in PRC2 binding at all normally DNA methylated regions could lead to the overall increase in the K27me3 modification globally observed in *Dnmt1*$^{-/-}$ MEFs. Alternatively, increased K27me3 binding may only be observed at certain regions that are normally DNA methylated and also compatible somehow with PRC2 binding. Many studies have now found negative correlations between the locations of DNA methylation and K27me3 across the genome, consistent with a high level of DNA methylation being inhibitory to PRC2 binding (Wu *et al.*, 2010; Lister *et al.*, 2009; Hawkins *et al.*, 2010; Lindroth *et al.*, 2008). This idea is also supported by a recent study that suggests DNA methylation of CpG-rich regions can prevent PRC2 targeting in mouse ES cells (Lynch *et al.*, 2011). The other side to this working model is the reduced K27me3 observed in *Dnmt1*$^{-/-}$ MEFs at normally K27me3 marked promoters. There have been several links drawn recently between PRC2 targeting in mammals and CpG islands (Branciamore *et al.*, 2010; Mendenhall *et al.*, 2010). It has been suggested that the default state of a CpG island devoid of activating transcription factor binding sites in mouse ES cells, and possibly in differentiated cells, is PRC2 binding (Mendenhall *et al.*, 2010; Lynch *et al.*, 2011). This was concluded as exogenous CpG-rich DNA that was inserted into these cells recruited the PRC2 complex and the K27me3 mark (Mendenhall *et al.*, 2010; Lynch *et al.*, 2011). A role for CpG islands in PRC2 recruitment is supported by the observation that Ezh2 associates with GC-rich DNA in mouse ES cells and

that many K27me3 marked regions are associated with one or more CpG islands (Tanay *et al.*, 2007; Ku *et al.*, 2008; Mendenhall *et al.*, 2010; Branciamore *et al.*, 2010). My data suggests that K27me3 is lost in *Dnmt1*$^{-/-}$ MEFs at promoters that contain a CpG island with low level DNA methylation. My working model predicts that this effect may arise due to a titration of the PRC2 complex from normal binding sites by increased binding throughout the bulk genome. This working model can be tested with the experiments proposed in the future work section of this discussion.

It is possible that the mechanisms proposed in chapter 5, involving intergenic transcription, altered CTCF binding and altered MLL occupancy, also contribute to the redistribution of K27me3 at some loci in *Dnmt1*$^{-/-}$ MEFs. These mechanisms were discussed in the discussion section of chapter 5 and will not be repeated here. Are MeCPs required as interpreters of the 5mCpG signal in order to properly target K27me3? The mouse phenotypes of knockout of MeCPs studied so far are not compatible with gross changes in K27me3 distribution or PRC2 target mis-expression where severe embryonic phenotypes would be expected (for example, Guy *et al.*, 2001; Zhao *et al.*, 2003). However, a role for MeCPs cannot be formally excluded as the possibility of redundancy among MeCPs has not yet been comprehensively excluded. If MeCPs are not involved then alternative interpretation of 5mC may be required. This could come directly from a core PRC2 component or an accessory protein. An *in vitro* assembled PRC2 complex consisting of only Ezh2, Eed, Suz12, RbAp48 and Aebp2 has been suggested to be inhibited in its binding to *in vitro* assembled chromatin by DNA methylation (Wu *et al.*, 2010). This suggests that one or more PRC2 components used in this assay are directly inhibited by DNA methylation. A candidate for the negative effect of DNA methylation is Aebp2 which interacts with PRC2 components and contains a DNA binding zinc finger domain (Cao and Zhang, 2004; Kim *et al.*, 2009b). It has not yet been reported whether the Aebp2-DNA interaction is influenced by DNA methylation. Using homologene (NCBI), a homolog of Aebp2 does not seem to exist in *Drosophila* and *C. elegans*, although this would have to confirmed using more detailed analysis. It would be interesting to perform *in vitro* chromatin binding experiments, with and without DNA methylation, using different components of the PRC2 complex, such as with and without Aebp2, to find out which proteins are important for this effect. Also, examining the effect of Aebp2 depletion on PcG targeting *in vivo* would be interesting as a mouse lacking Aebp2 has not yet been generated. It will be interesting to find out more about the apparent inhibition of PRC2 binding by DNA methylation, in particular by studying the proteins and mechanisms of inhibition that are involved.

The results presented in this thesis suggest that the DNA methylation system may play a major role in targeting PRC2 and the K27me3 mark to specific regions of chromatin. Support for this idea in the literature comes from studies investigating K27me3 in DNA hypomethylated mouse mutant systems. Interestingly, treatment of the human cancer cell line, 293T, with demethylating agent 5-azacytidine (a similar demethylating agent to 5-aza-dC) resulted in loss of K27me3 from a large number of gene promoters that before treatment were highly marked by this modification (Komashko *et al.*, 2010). Also in this study, the global levels of the K27me3 mark were shown to be increased by western blot in cells treated with 5-azacytidine (Komashko *et al.*, 2010). Although this study uses cancer cells, which would be expected to have a drastically different DNA methylation profile from normal cells, it provides an encouraging independent validation of the observations made in this thesis. Reduced K27me3 has been reported at a small number of PcG target genes in *Lsh*$^{-/-}$ MEFs, which exhibit reduced DNA methylation at these regions (Xi *et al.*, 2007; Tao *et al.*, 2011). Also, cells from ICF patients carrying *DNMT3B* mutations display reduced K27me3 and PRC2 association at a small number of assayed gene promoter regions when compared to controls (Jin *et al.*, 2008). In this study, small decreases in DNA methylation were observed at these promoters but, of note, they are suggested to contain very little DNA methylation in control cells (Jin *et al.*, 2008). This appears similar to my observations in *Dnmt1*$^{-/-}$ MEFs, where gene promoters that show the greatest loss of K27me3 are associated with CpG island promoters with apparent low level DNA methylation normally. Reduced DNA methylation has also been suggested to lead to increased K27me3 at certain regions. Knockout of *Dnmt3a* in mouse neural stem cells reportedly results in decreased expression of many genes required for neurogenesis (Wu *et al.*, 2010). Here, decreased DNA methylation from regions flanking promoter CpG islands was suggested to lead to increased binding of the PRC2 complex, increased K27me3 and inhibition of transcription (Wu *et al.*, 2010). In support of this idea, the PRC2 complex was shown to be inhibited by DNA methylation in an *in vitro* chromatin methylation assay (Wu *et al.*, 2010). In this study, many gene promoters also showed decreased K27me3 and were transcriptionally upregulated in *Dnmt3a* mutant neural stem cells, but this was not reported in detail as it was not a focus of the paper (Wu *et al.*, 2010). In addition, western blot for K27me3 in *Dnmt3a* mutant neural stem cells hinted that overall levels of this modification may be slightly increased relative to wildtype cells (Wu *et al.*, 2010). As a form of validation, it will be interesting in the future to analyse the data from Wu *et al.* (2010) in more detail in respect of K27me3 changes, particularly in at the promoters found in my study to lose K27me3 in *Dnmt1*$^{-/-}$ MEFs. In a recent paper, DNA hypomethylated *Dnmt3a;Dnmt3b*-deficient ES cells

were shown to exhibit *de novo* marking of normally-methylated CpG-rich regions with K27me3, supporting this idea (Lynch *et al.,* 2011). Together, these studies support my data suggesting for a role of DNA methylation in correct targeting of the PRC2 complex to chromatin. Importantly, these data are consistent with my working model where widespread loss of DNA methylation leads to redistribution of K27me3 patterns, from PRC2 targets that are normally marked by low level DNA methylation, to non-PRC2 targets that are normally marked by high level DNA methylation (Fig. 6.1).

6.2. A potential new role for DNA methylation: Implications for gene regulation in development and disease

What do these observations tell us about the roles of DNA methylation in the cell? Primarily, they suggest that an intact DNA methylation system is required for the correct distribution of the PRC2-associated K27me3 histone mark. If true, this suggestion supports a major role for DNA methylation in gene regulation out with of canonical promoter DNA methylation-mediated gene repression. Crucially, it means that genes associated with an unmethylated CpG island at their 5' end can still be dependent on DNA methylation for repression. This considerably widens the scope for genes that can be regulated by DNA methylation, to include many highly tissue-specific genes with crucial roles in embryonic development that would not classically be associated with regulation by DNA methylation. This could help to explain the relatively large number of genes that are reportedly upregulated when *Dnmt1* is mutated in somatic cells, despite the relatively small number of genes that are thought to be associated with DNA methylation at their promoter regions (Jackson-Grusby *et al.,* 2001; Lande-Diner *et al.,* 2007; Shen *et al.,* 2007; Illingworth *et al.,* 2008; Meissner *et al.,* 2008). It could also contribute to an explanation for the phenotype of severe *Dnmt1* mutation in mouse, where development is arrested and several structures fail to form (Lei *et al.,* 1996). In support of a role of DNA methylation in the regulation of PcG target genes, *Hox* gene mis-expression has previously been associated with DNA hypomethylation in multiple mouse mutant contexts, but the reason for this effect has remained unclear (Xi *et al.,* 2007; Tao *et al.,* 2010; Velasco *et al.,* 2010). An example is upregulation of certain *HoxA* genes in mice carrying compound heterozygous mutations in *Dnmt3b*, which occurs together with a homeotic skeletal transformation that is characteristic of *Hox* gene mis-expression (Velasco *et al.,* 2010). These observations could potentially be explained by a role for DNA methylation in PRC2 targeting to *Hox* clusters, as oppose to a

canonical promoter repression role of DNA methylation. It will be interesting to investigate the cause of *Hox* gene mis-expression in more detail when DNA methylation is altered, particularly to try to dissect different roles that DNA methylation may play at different loci.

If true, this new role for DNA methylation could have implications for aberrant gene regulation in disease. It has been observed that many genes found to be mis-regulated in patients with ICF syndrome, caused by *DNMT3B* mutations, are associated with CpG island promoters with low level DNA methylation in normal cells (Jin *et al.,* 2008). Furthermore, many of these genes are PRC2 targets and appear to have reduced K27me3 and association with the PRC2 component Suz12 in ICF patient cells, correlating with increased transcription (Jin *et al.,* 2008). Mis-expression of some of these genes, with their roles in neurogenesis, embryonic development and immune function, is consistent with some of the phenotypes of ICF syndrome. Here, PRC2 targeting may be directly influenced by altered DNA methylation patterns in ICF patients, providing an explanation for mis-expression of these genes that would not normally be connected with regulation by DNA methylation. PcG target genes often play central roles in cell fate decisions, and, repression of lineage-specific genes in multipotent progenitors by PcG proteins is thought to be crucial for maintenance of the stem cell state (Boyer *et al.,* 2006; Bracken *et al.,* 2010). Recently, *Dnmt1* was shown to be required for maintaining adult mouse stem cell populations in the multipotent state (Broske *et al.,* 2009; Trowbridge *et al.,* 2009; Sen *et al.,* 2010). Specifically, depletion of Dnmt1 in hematopoietic stem cells and epidermal progenitors resulted in premature differentiation and differentiation skews towards certain cell lineages (Broske *et al.,* 2009; Trowbridge *et al.,* 2009; Sen *et al.,* 2010). Although loss of canonical promoter DNA methylation-mediated repression may account for some of the observed effects, it is interesting to speculate that an effect on PcG target genes may contribute to these phenotypes. Altered DNA methylation patterns have also been linked to certain leukaemias. For example, mutations in *DNMT3A* and *TET2* have both been associated with certain types of leukaemia, where DNA methylation changes have been reported (Ley *et al.,* 2010; Yan *et al.,* 2011; Ko *et al.,* 2010; Ko *et al.,* 2011). One report suggests that DNA methylation patterns are altered at *Hox* genes in patients with *DNMT3A* mutations and that many *Hox* genes are differentially expressed (Yan *et al.,* 2011). It is interesting to speculate here that K27me3 changes are driving expression changes at these genes as a consequence of DNA methylation changes. Redistribution of DNA methylation patterns have been reported in a variety of cancers and in some cases are thought to contribute to the cancer phenotype (Irizarry *et al.,* 2009; Chen *et al.,* 2009). If my working model is correct then these changes would be expected to cause rearrangement of PRC2 binding and the K27me3 mark in cancer.

As PRC2 and K27me3 are well associated with the regulation of transcription, this could contribute to the aberrant epigenome that is observed in cancer (Boyer *et al.,* 2006; Bracken *et al.*, 2006; Wu *et al.*, 2010).

These findings are also interesting from a polycomb research standpoint. How PcG proteins are targeted to specific regions of the genome is not well understood and represents a major question in this area of research. There is a growing connection in the literature between the DNA methylation and polycomb systems but exactly how they are related is unclear. Much of the recent emphasis has been on a putative role for the PRC2 complex in targeting DNA methylation, either during cellular differentiation or in cancer (Mohn *et al.*, 2008; Vire *et al.*, 2006). However, epigenomic studies have suggested that the distribution of DNA methylation and K27me3 marks are in fact negatively related suggesting that this mechanism is not widespread (Lister *et al.*, 2009; Hawkins *et al.*, 2010). The data presented in this thesis add to the growing connection between the two epigenetic systems, and further suggest that an extensive interplay exists between them. Specifically, it is becoming apparent that the DNA methylation mark may act as a major factor in determining PRC2 binding throughout the genome. A better comprehension of this process will be crucial to understanding how PcG complexes are targeted in mammals.

6.3. How is PRC2 targeted in organisms lacking CpG methylation?

An important evolutionary question is raised by this work. Assuming CpG methylation does play a major role in targeting of the PRC2 complex and the K27me3 mark in mouse cells, how is PRC2 targeted in organisms lacking CpG methylation? *Drosophila melanogaster* and *Caenorhabditis elegans* each contain no CpG methylation yet utilise PcG complexes in transcriptional repression. It is therefore interesting to consider PRC2 targeting in these organisms in order to understand how a role for DNA methylation in this process could have evolved. In *C. elegans*, although PcG homologs have been implicated in gene regulation and cell fate decisions, relatively little is known about the PcG system (Yuzyuk *et al.*, 2009). In comparison, much of the early work on PcG proteins was performed using *Drosophila* as a model organism and it is well established that PcG proteins are responsible for repression of many genes, including *Hox* gene orthologs, that have major roles in embryonic development (Schwarz and Pirrotta, 2007). In fact, the mammalian orthologs of many of these developmental regulators are also PcG targets, suggesting a distant conservation of regulatory mechanisms at these genes (Schwarz and Pirrotta, 2007). The *Drosophila* genome does not encode homologs of mammalian Dnmt1 or Dnmt3 proteins and, although

the extent and location of DNA methylation in this organism is a subject of debate, it is clear that DNA methylation is not at all abundant and that CpG methylation is not present at all (Suzuki and Bird, 2008; Phalke *et al.,* 2009; Schaefer *et al.,* 2010b). For these reasons I will focus on *Drosophila* as an example of an organism with no CpG methylation. Why would processes so well conserved as targeting PRC2 to developmental genes differ between fly and mammals?

First of all it must be stated that our understanding of PRC2 targeting in any organism is limited and so any argument presented here is based on a certain amount of speculation. However, in *Drosophila* a link with sequence-specific DNA binding factors in PcG targeting is much better established than in mammals. In particular, the *Drosophila* Pho protein is bound to PcG repressed regions and may be involved in targeting PcG complexes (Schuettengruber *et al.,* 2009). DNA elements, known as polycomb response elements (PREs), are capable of recruiting PcG complexes and inducing stable repression of adjacent reporter genes in *Drosophila* (Cavalli and Paro, 1998; Cavalli and Paro, 1999). The search for PREs in mammalian genomes has so far produced few results. A couple of PRE-like elements have been reported in mammals which have some of the characteristics of fly PREs (Sing *et al.,* 2009; Woo *et al.,* 2010). The closest mammalian protein to a homolog of fly Pho is YY1. YY1 has been implicated in activity of the PRE-like element described in Woo *et al.* (2010). However, a wider role of YY1 in PRC2 targeting in mammals seems unlikely as a clear overlap between YY1 binding motifs and PRC2 binding sites has not been observed (Xi *et al.,* 2007; Margueron and Reinberg, 2011). In fact YY1 binding motifs are enriched in CpG islands that are not bound by PRC2 complex components in mouse ES cells (Ku *et al.,* 2008). There is also a difference in the pattern of PcG protein binding between fly and mammals. In fly, all PRC2 and PRC1 components that have been mapped so far, with the exception of Pc, bind in sharp peaks at PREs and the K27me3 mark is found spread from these sites in large blocks (Schuettengruber *et al.,* 2009). In mammals, PRC2 components are bound in a pattern more similar to the K27me3 mark, in regions covering small or large domains of chromatin, but not in sharp peaks (Boyer *et al.,* 2006; Lee *et al.,* 2006; Bracken *et al.,* 2006). These apparent differences suggest that targeting mechanisms may be different between fly and mammals (Schuettengruber and Cavalli, 2009). Perhaps in fly, PRC2 targeting is more reliant on PRE-like mechanisms where the proteins are recruited to defined genomic loci from which the K27me3 mark can spread.

The only DNA sequences that have been convincingly associated with PRC2 binding sites in mammals so far is CpG islands (Tanay *et al.,* 2007; Ku *et al.,* 2008; Mendenhall *et al.,* 2010). Certain CpG islands that represent targets of PRC2 binding in ES

cells show evidence for strong conservation of CpG density suggesting that these regions are important, possibly for PRC2 recruitment (Tanay *et al.*, 2007; Branciamore *et al.*, 2010). CpG depletion from the bulk genome is thought to have occurred mainly via spontaneous deamination of 5mC to uracil and subsequent failure to repair (Bird *et al.*, 1985). It should also be noted that the GC content of CpG islands is higher than the rest of the genome, which cannot be explained by the increased CpG content alone (Bird, 2011). The *Drosophila* genome, with no CpG methylation, does not show CpG depletion and so relative to vertebrate genomes, in some ways is like one big CpG island (Bird, 2011). Also, the *Drosophila* genome does not show evidence for selection of high CpG content within *Hox* gene coding regions that is observed in vertebrate genomes (Branciamore *et al.*, 2010). It is tempting to speculate that binding of the PRC2 complex in *Drosophila* is potentially more permissive, whereas in vertebrates, binding may be guided by CpG islands. CpG islands may be preferred binding sites of the PRC2 complex in vertebrates for a number of reasons, such as their low DNA methylation, high GC content or another aspect of their unique chromatin state. As the *Drosophila* genome contains no CpG methylation and displays no depletion of CpG dinucleotides or GC content it may be more compatible with PRC2 binding. Perhaps for this reason PRC2 recruitment in *Drosophila* is more reliant on a PRE-like mechanism as they have lost CpG methylation during their evolution. Evidence for PRC2 inhibition by DNA methylation in distantly related species would be very interesting here. Interestingly, correlative evidence for this process in *Arabidopsis* has been reported, where the presence of DNA methylation in vegetative tissues at genes that are PRC2 homolog targets in endosperm, correlates with lack of PRC2 binding in vegetative tissues (Weinhofer *et al.*, 2010).

6.4. Future work

In terms of immediate future work I think it will be important to map DNA methylation in primary cells and tissues with special attention paid to the *Hox* clusters. This will be important to understand the reason for their mis-expression in DNA hypomethylated cells. In order to test the working model of K27me3 redistribution in hypomethylated cells (Fig. 6.1), it will be crucial to determine where increases in the K27me3 mark have occurred in *Dnmt1*$^{-/-}$ MEFs, and how these changes relate to DNA methylation status. As K27me3 increase appears to be rare at gene promoter regions, a more genome wide approach here is required. ChIP-chip using genome wide tiled microarrays or ChIP-seq (ChIP followed by sequencing) for the K27me3 mark would be useful here as regions outside promoters can be

studied. Studying the characteristics of regions that gain K27me3 in *Dnmt1*$^{-/-}$ MEFs, including their DNA methylation status in normal and *Dnmt1*$^{-/-}$ cells, will hopefully provide key information on the processes driving K27me3 redistribution in these cells.

To ask if the DNA methylation mark itself is important it will be interesting to attempt to rescue the defects in K27me3 targeting in *Dnmt1*$^{-/-}$ MEFs by reintroducing wildtype and catalytic inactive Dnmt1 proteins. This is an experiment that I have been attempting for a while but so far have been unsuccessful. A rescue with wildtype Dnmt1 would also help to exclude that changes are due to an artefact of the *Dnmt1*$^{-/-}$ MEF cell line. However, it should be noted that even if Dnmt1 protein levels are re-established in these cells, it is not known if DNA methylation will be fully restored. Rescue of DNA methylation levels has been achieved by Dnmt1 expression in ES cells deficient for *Dnmt1* (Damelin and Bestor, 2007; Takebayashi *et al.*, 2007). ES cells appear to possess a high level of Dnmt3 proteins and *de novo* methyltransferase activity, perhaps allowing them to be able to re-establish methylation patterns (Xie *et al.*, 1999; Okano *et al.*, 1998a). It remains to be seen if the same will be true for MEFs where *de novo* methyltransferase levels are expected to be lower.

Importantly, multiple experimental systems were used during this project, suggesting against the possibility that differences in gene expression and histone modifications observed in *Dnmt1*$^{-/-}$ MEFs represent artefacts of this cell line. However, I think that this study would benefit from a more extensive validation of the observations using another DNA hypomethylated system. For example, a more detailed analysis of K27me3 distribution in cells treated with 5-aza-dC would help to confirm observations made using *Dnmt1*$^{-/-}$ MEFs. It may be interesting to investigate ES cells lacking all three active DNMTs as these cells are nearly completely lacking CpG methylation (Tsumura *et al.*, 2006). Knockout of different combinations of DNA methyltransferases in ES cells results in different patterns of hypomethylation potentially allowing dissection of the effect of DNA methylation on PRC2 targeting (Jackson *et al.*, 2004).

In the longer term, ideally, I would like to use a more physiologically relevant system for DNA hypomethylation, to confirm the observations made using *in vitro* cell models. This could include more work using *Dnmt1* mutant embryos. The use of conditional knockout cell lines established from mouse embryos, or defined cell populations such as haematopoietic stem cells could prove useful here and provide a more physiological angle to the study. It may also be interesting to examine the effect of DNA methylation on PRC2 targeting in other organisms. Distant vertebrates, such as zebrafish or *Xenopus laevis*, could be used here to study this effect in other organisms with global methylation patterns,

and provide information about conservation of this mechanism. Finally, I think it would be interesting to examine the effect of an altered DNA methylome in cancer, on the PRC2 complex and K27me3 mark. Following my model, as DNA methylation patterns are drastically redistributed in certain cancers, this would be expected to cause substantial changes in binding of the PRC2 complex to chromatin, and hence contribute to the 'cancer epigenome'.

Appendix 1 - Summary of custom tiling array design

Locus	Length of tiled region (bp)	Coordinates in UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly
Hoxa	818,585	chr6:51,781,416-52,600,000
Hoxb	755,001	chr11:95,845,224-96,600,224
Hoxc	702,398	chr15:102,400,000-103,102,397
Hoxd	668,990	chr2:74,241,352-74,910,341
Hba	237,547	chr11:32,087,454-32,325,000
Hbb	197,639	chr7:110,861,362-111,059,000
Shh locus	1,221,001	chr5:28,489,000-29,710,000
Kcnq1 locus	394,001	chr7:150,256,000-150,650,000
Dlk1 Meg locus	136,468	chr12:110,693,533-110,830,000
Gnas locus	69,619	chr2:174,105,382-174,175,000
H19 locus	137,587	chr7:149,730,414-149,868,000
Mest-Kl14 locus	225,001	chr6:30,685,000-30,910,000
Pax6	35,000	chr2:105,505,001-105,540,000
Brachyury	150,001	chr17:8,554,691-8,704,691
Gsc	50,001	chr12:105,687,160-105,737,160
Hhex	40,001	chr19:37489346-37529346
Nestin	28,077	chr3:87,765,656-87,793,732
Lin28	27,001	chr4:133,553,000-133,580,000
Pou5f1	9,401	chr17:35,640,600-35,650,000
Sox2	23,606	chr3:34,540,712-34,564,317
Nanog	10,001	chr6:122,656,000-122,666,000
Klf4	13,200	chr4:55,536,078-55,549,277
C-Myc	15,063	chr15:61,811,875-61,826,937
Mirlet7-D	14,164	chr13:48,630,830-48,644,993
Gata1	17,112	chrX:7,532,108-7,549,219
Gata4	13,904	chr14:63,855,878-63,869,781
Cdx2	12,700	chr5:148,110,571-148,123,270
Asl1/Mash1	6,000	chr10:86,953,114-86,959,113
Nkx2-9	6,000	chr12:57,709,722-57,715,721
Myf5	6,262	chr10:106,919,672-106,925,933
Olig2	7,038	chr16:91,223,137-91,230,174
Magea3	8,801	chrX:151,379,200-151,388,000
Tex19.1	8,770	chr11:121,004,657-121,013,426
Dnmt3b	16,799	chr2:153,470,702-153,487,500
Rpl19	6,164	chr11:97,886,000-97,892,163
Actb	5,414	chr5:143,663,892-143,669,305
Aard	8,426	chr15:51,869,406-51,877,831
Adssl1	31,878	chr12:113,852,987-113,884,864
Ankhd1	6,329	chr18:36,715,293-36,721,621

Aurkb	9,030	chr11:68,856,737-68,865,766
Brd3	41,055	chr2:27,296,495-27,337,549
Chchd7	10,373	chr4:3,862,488-3,872,860
Chmp4b	56,637	chr2:154,473,323-154,529,959
Crabp1	12,453	chr9:54,610,540-54,622,992
Ctnnal1	81,381	chr4:56,810,244-56,891,624
Ddx6	48,314	chr9:44,404,928-44,453,241
Emilin2	88,185	chr17:71,586,819-71,675,003
Eno1	8,210	chr4:149,609,995-149,618,204
Fgf10	115,623	chr13:119,484,571-119,600,193
Fus	21,830	chr7:135,107,354-135,129,183
Gypa	25,112	chr8:83,013,758-83,038,869
Hebp1	46,046	chr6:135,079,862-135,125,907
Hemgn	15,293	chr4:46,404,312-46,419,604
Hnrpa1	8,303	chr15:103,068,649-103,076,951
Jarid1b	8,076	chr1:136,453,744-136,461,819
Ly6a	6,777	chr15:74,823,425-74,830,201
Ly6c1	5,724	chr15:74,874,491-74,880,214
Mapk1ip1l	37,167	chr14:47,911,795-47,948,961
Mlkl	39,156	chr8:113,829,174-113,868,329
Npm3	5,490	chr19:45,820,805-45,826,294
Nrn1	13,244	chr13:36,815,287-36,828,530
Pcdha7 locus	1,012,030	chr18:36,996,988-38,009,017
Pdlim2	20,183	chr14:70,560,661-70,580,843
Ppm1b	5,838	chr17:85,352,959-85,358,796
Ppp1cc	5,664	chr5:122,604,605-122,610,268
Pramel4	15,290	chr4:143,646,480-143,661,769
Prf1	9,716	chr10:60,757,991-60,767,706
Psip1	36,768	chr4:83,097,979-83,134,746
Shmt2	11,919	chr10:126,950,487-126,962,405
Sif1	13,637	chr16:18,262,780-18,276,416
Slc7a3	9,198	chrX:98,273,946-98,283,143
Snx17	12,596	chr5:31,492,177-31,504,772
Srebf2	86,538	chr15:81,963,276-82,049,813
Tcfap2a	27,278	chr13:40,807,558-40,834,835
Tle1	4,134	chr4:71,858,971-71,863,104
Tnik	6,790	chr3:28,159,250-28,166,039
Xlr4a	12,245	chrX:70,317,643-70,329,887
Zbtb2	31,535	chr10:5,953,177-5,984,711
Xist	34,292	chrX:100,649,996-100,684,287
Total	**8,200,080**	

Appendix 1 - Summary of custom tiled Nimblegen microarray regions.

Appendix 2 - Gene lists defined in chapter 4 through epigenomics

Lists of gene promoters defined in chapter 4 as having significantly different enrichment of either K27me3 or K4me3 in *Dnmt1*$^{-/-}$ relative to *Dnmt1*$^{+/+}$ MEFs as measured by promoter microarray. Changes at these promoters surpasses thresholds set according to chapter 2.

K27me3 lost in *Dnmt1*$^{-/-}$ MEFs (difference < -0.6815459 (for *Dnmt1*$^{-/-}$ - *Dnmt1*$^{+/+}$))

Hoxa7	Hoxc6	Hoxa11as	Hoxc5	Hoxc11	Hoxc10
Hoxc9	Vax2	Zic5	Hoxc12	Hoxa9	Mir10a
Hoxb6	2700086A05Rik	Hoxa10	Mecom	Hoxc8	Psmb8
Insl3	Hoxd12	Hoxa4	Hoxb5	Hand2	Rpp25
Hoxb2	Hoxc4	Hoxa6	Tnfsf9	Hoxb3	Pax9
Pde1b	Tbr1	Hoxa13	Mfsd7a	Tcfap2a	Nr4a3
Esx1	Pitx1	Bhlhe22	Trim14	Gramd1b	Mkx
Foxf2	Gatm	Shank3	Grtp1	Nmnat2	Dmrta2
Hoxc13	Hoxd4	Hand1	Neurog3	Fam113b	Ripk4
Hoxb7	Cldn5	Dnm3	Prdm8	Gm5089	Foxq1
Syt13	AI118078	Mfsd6	Gata4	Has3	Csf2rb2
Pax6	Dnaja4	Csf2rb	Slc6a5	Rem1	Crlf1
Frzb	Btnl7	Fam69c	Gfra4	EG547347	Srgn
Hoxd10	Ddx60	Osr2	Msx1	Uts2r	Wnt2b
Wnt10a	Kcnc2	Macf1	Zbtb8b	Wnt5b	Clic5
Tacc1	Nrgn	Adam33	Gata2	Kcnk5	Atp6v1b1
Fbxo39	Itpr1	Ptpn5	AI182371	Mapk15	Hoxd9
Aim1	Ebf2	Dmrt2	Nos3	Fgf17	Tcf21
B3gnt3	Mppe1	Ppp1r3c	Isl1	T	Tap1
Slc16a12	Lmx1a	Smoc2	Gpam	Spink2	Panx1
Rap1gap	Phox2b	Tbx4	Slc47a2	Fli1	Sox14
Evx1	Foxd4	Tap2	Tll2	Foxa1	Spatc1
Cck	6430571L13Rik	Sim2	H2afy2	Usp18	H2-D1
Myoc	Gata6	Parp14	Il17d	Fgf8	Crabp1
Prss16	Kcna1	Selp	Slmo1	Apc2	Gdf7
Col13a1	Hspa12a	Foxb2	Sele	Chn2	Fabp3
Tmed8	Cxxc5	Nr5a2	Clec11a	Alx1	Phldb3
Dpep1	Dscam	Prox1	Foxb1	Sv2c	Raet1e
Map3k9	Sp5	Gdnf	Sp9	Frmd5	Vipr1
Shisa7	Esrra	Lbx2	Inmt	Tbx20	Enpp2
Lhx2	Neurod1	Notum	Prdm1	Ifit1	Strc

Grhl2	Cldn3	Ramp1	Arg2	Olig3	Fgd4
Tbx5	Nkx2-3	Adra1a	Egflam	1500009L16Rik	Epb4.1l4a
Nrg4	Cabp2	Slc13a5	Isl2	Dlx4	Fbln7
Herc3	Podxl	Zfp385a	Arhgap28	Hhipl1	Adrb1
Hoxa2	Rgl3	Fam40b	Fam83h	Ppp1r1a	C530028O21Rik
Rhou	Gfi1	Gal	Afap1l2	Pigz	Atp2b3
Kcnq4	Cyp1b1	Acpl2	Tmem221	Ppp1r3b	Sell
Sim1	Fam89a	Fezf2	F11r	Tcp11l1	Eya1
Wnt3	Irx6	Mgat4a	Six3	Spats2l	Ebf1
Nphs2	Trpc1	Sectm1a	Isoc2b	Prlhr	Ttc7b
Slco4a1	Htr6	0610010O12Rik	Islr2	Cyp26a1	Gsx2
Zic2	Ezr	Zfp503	Pdgfra	9030224M15Rik	Nkain4
Cilp2	Lrtm1	Greb1	Xlr4a	Peli2	Cyp27a1
Gbp6	Hnf1b	Arhgap9	Phactr2	Dcst1	Kcns2
Ttc9	Fam149a	Neurog1	Bmyc	Usp44	Nkx6-3
Vax1	Fmo2	Nxph4	Helt	9530036O11Rik	Otop1
Sema4g	Rgs20	Aldh1l1	Gm5878	Cdk5r2	AA388235
Tbx2	Pla2g3	Ngb	Kcnh2	Rax	Slc16a13
Otx2	Hoxd3	Slc6a2	Ddn	Prdm6	Pdx1
Wnt9b	Pls1	Raet1a	Hmx1	Cryba2	Hmx2
Sfmbt2	Dll4	Ido1	Klf1	Prr18	Nkx2-2
Gm266	Nkx1-2	9030612E09Rik	Rln1	Higd1c	Tprg
Six6	2010015L04Rik	Kcnk13	Creg2	BC089491	Antxr1
Ptpn7	Ltb4r1	Cyb561	Nr2e1	Hpse2	Pax7
D930020E02Rik	Kis2	Dlx6os1	Gata5	Gpr162	Tcf7
Enpp1	Il13	Nefl	Anpep	Kdelr1	Arx
Fxyd6	Prdm12	Nkx6-2	Srd5a2	Adora1	Arl5c
Clec1b	Myod1	Rgs10	Selm	Itga11	Hoxd11
Gkn1	Slc6a4	Dpys	Nckap1l	Nppc	Ccno
Nrarp	Icam2	Gm3238	Rbpms	5730409E04Rik	Dbx1
Mlkl	Foxa2	F5	Dmrt3	Stc2	Ckmt1
Hspa12b	Sall1	Msx3	Tcfap2c	Zbtb16	Raet1d
Tmprss7	Ptpn22	Sox1	Gstt2	Sp8	Tcfcp2l1
Slc32a1	Ptprf	Igsf3	Slc27a2	Sectm1b	2900005J15Rik
Spock2	Msx2	Atoh1	1200009O22Rik	Rbpjl	Aqp3
Col27a1	Cd27	Kcnip2	Zar1	Pdlim3	Slc47a1
Tnfrsf19	Rgs11	Rfx4	Klhl26	Pip5k1b	Lgr6
Pigh	Kcnh5	Cybrd1	Lmbr1	Lyn	Trim36
Sigirr	Gcnt1	Cd24a	Sall4	Apoh	Barhl1
Tmem181a	Ptp4a3	Stag3	Acta1	Pax5	Spats1
Prok2	Parvb	Wnt2	Rapgef4	Cxcr4	Atp1b2
7420416P09Rik	Ankrd43	Pla2g10	Nkx2-5	Sel1l3	C1ql2
Il24	Calcb	H2-K1	Olig2	Ica1l	Nags
Egfr	Hc	Pacsin3	Gcgr	Nkx3-2	Lrp2
Chst11	Rasl12	Smtnl2	Pou4f2	Rtbdn	Zdhhc23

Barx2	Celf6	Dock8	Kcp	H2-Bl	Rab19
Tal2	Mpzl2	Gm2002	BC068157	A4galt	Ptf1a
Osbpl5	Slc9a2	Hs3st3a1	Ankrd56	Abcb9	Ntng2
Kremen2	Tmem210	Plcb2	Ism1	Syt6	Ugt1a7c
Calcr	Rasgrp2	Sncaip	Npy	H2-Q1	Lrrc10b
Kcnip3	Adcy1	Wfdc2	Gad2	Tlx1	H2-Tw3
Hpcal4	Otp	Mafb	Apobec3	Icmt	Nfasc
Cgn	Tex19.1	Lhx8	Bub3	Pcdhac2	Vipr2
Shmt2	Gng4	Ppp2r2c	Fam189a2	Ugt1a2	Rnf32
Tnfrsf10b	Flt4	Penk	Lhx5	Hoxb13	Gdf6
Kcnk4	Cdh1	Crip1	Abcg8	Eps8l1	Adipor2
Nhlh2	Rab15	Kcna6	Prkg1	Mnx1	Slc16a9
Gm12824	Gpr3	Prrxl1	Jak3	Frmd4a	Mixl1
Slc6a20a	Rab17	Evx2	Nkx2-6	1110032A04Rik	Disc1
Robo2	Barx1	G630025P09Rik	Bhlha9	Cacng6	Gal3st4
Fmo5	Gm10565	Lonrf3	Dhh	Wt1	Apol10b
Spocd1	Htr2a	Tubb4	Duoxa2	Klf15	Kcnh8
Prdm16	Prr15	Gbp5	Tyro3	Cenpv	Krtap5-1
Kbtbd12	Tcfap2d	Prdm13	Irf5	Pcnxl2	Lrrn2
Hapln4	Scnn1a	2410137M14Rik	6430704M03Rik	Depdc7	AI661453
Hdc	Nr1h3	Il28ra	Prdm14	Esrp1	Hpse
Sox7	Gstm4	Npm2	Tbx1	Dnajc22	Cpne7
Gap43	Cntn2	Olig1			

K27me3 gain in *Dnmt1*$^{-/-}$ MEFs (difference > 0.5656917 (for *Dnmt1*$^{-/-}$ - *Dnmt1*$^{+/+}$))

Olfr648	AU015228	Ankrd7	Acot10	Snord116	Olfr461
Gm15085	Iltifb	Pcdhb20	C130026I21Rik	Il23r	H19
Olfr272	Rex2	Gm428	Sp140	Rit2	Tmem176a
Acpp	Cyp7a1	Myh1	EG381936	Dbp	Tmem125
Prr23a	Il22	Pcdhb18	Gm13051	Slc1a3	2610305D13Rik
Sftpa1	Olfr649	A530032D15Rik	Gm15127	2610528J11Rik	Myh4
Rbmy1a1	Darc	Gm10922	Gm3750	Gm2863	

K4me3 gain in *Dnmt1*$^{-/-}$ MEFs (difference > 0.5363075 (for *Dnmt1*$^{-/-}$ - *Dnmt1*$^{+/+}$))

Hoxc10	Hoxa11as	Hoxc11	Hoxa9	Hoxa10	Nostrin
Xdh	Rbmy1a1	Csf2rb2	Cort	Hoxc9	Mfsd7a
Agrn	Hk3	Arhgap9	Csf2rb	Rltpr	Tuba3b
Rhox4a	Fyb	F10	Lat	Endou	Irf7

Arl9	Gm5105	Spink2	Oplah	Olfr282	Pou6f1
Sipa1l2	Clec11a	Gpat2	Anpep	Jph2	Rho
Wfdc15a	Tnfsf13b	Tex13	4930511M11Rik	Hand2	Spatc1
Efnb3	Hoxd10	Mia2	Amn	Tlr6	Kcnu1
Thbs2	A430078G23Rik	Gjb5	Adam33	Hnf1b	Ido1
Tex19.1	Nlrp4c	Sectm1b	Hoxc12	Rhox4e	Sdr42e1
Dnajb8	Actl7a	Wnt10a	Chad	Rnf186	Slfn3
Nat2	Cela3b	1700018B24Rik	Olfr750	Dpep2	Aldh1l1
Fam83h	Clec1a	Xlr4a	Hsh2d	Lrrc10	Hoxc13
Sla2	Sh3d19	Sycp3	Sstr2	Olfr234	Zfhx2as
Apob48r	Lim2	Krt18	Dennd1c	Spocd1	Pla2g5
4933403G14Rik	Esr1	Gbp5	Higd1c	Xlr3a	Slc15a3
2610036D13Rik	Col17a1	Thbd	2200002K05Rik	1700018L24Rik	Syce1l
Dcst1	Mylk3	Krt23	Aox4	Naprt1	1700039E15Rik
Kcnn4	Csl	1700090G07Rik	Sectm1a	Mb	Inhbe
Sgcg	Gm13011	Nphs2	Icam2	Gm1968	Aim2
Cabp2	Xlr3c	Mov10l1	Nrap	Mab21l1	Vax2
Hand1	Xlr3b	Dnase1l3	Apol10b	Klb	Gpr37l1
Ccdc135	Klhdc7b	Gm8884	Cacng6	Cxcl1	Lypd2
Dhx58	Lrrc25	4930563D23Rik	Nrl	Olfr71	Ido2
Rhox4f	Mpeg1	Cpsf4l	Jakmip1	Mboat4	Ly86
1700018B08Rik	LOC100039801	Clcnka	Ankrd36	Hoxd11	1700061G19Rik
Sparcl1	Mmp27	Catsper1	Tnfrsf14	Gm12776	Crispld2
Odf3l2	Gm7616	Hoxa7	Gtsf1l	Gzmk	Klk10
Wfdc1	Isoc2b	Naip6	Nat3	Krt36	Inpp5d
Pon3	Sirt5	Apob	Ccdc42	Sult2b1	Cmtm2a
Sohlh2	Rhox4b	Gm16378	Cd27	Dkk4	Anxa8
Tmem150b	Dpys	Slfn4	Nhedc1	Ghrh	Slc36a2
Gatm	Pi15	Apol7b	Acrbp	Gdf2	Rps4y2
Qrich2	Lepr	Mettl8	Phc2	Gstp2	Sult5a1
Cpa5	Cyp26c1	C1qtnf3	BC016201	Arhgef18	Tagap
Krt14	Prss38	Aqp6	Atp6v1e2	Rpl39l	Evx1
Blnk	Stra8	Tnf	Tmem232	Prkag3	C4a
Pla2g4f	Cdh5	Gdf5	Trpd52l3	Kcnj14	Oasl1
BC027072	Htatip2	Chrna2	Hoxa13	Wfdc10	Tfpi
Tdrd1	Slc13a5	Slc22a13	Zfp366	S100z	BC054059
Ranbp3l	Kcng1	Elfn2	Pcgf5	Kiss1	Batf2
1700094C09Rik	Car6	Hdc	Il12a	Itgb2	Pla2g2e
4930519G04Rik	Krt17	Naip5	Arl11	Lgals9	Rab37
Cmtm2b	Gm7534	Pkd1l3	Slc24a5	Hspb3	Tcte1
Bhlha15	Sgk2	Cdkl4	Gm1006	Nkx2-3	Sit1
Gm12250	Syce1	Apobec4	Pdzk1	Tm6sf2	Slco4a1
4930538K18Rik	Eps8l3	Msln	Dok2	6430548M08Rik	Cdc42ep2
Gramd1b	Itgb6	Zfp709	Gsdmc3	Actbl2	Adprhl1
Clec10a	Pnlip	Csad	Myod1	Sfpi1	Lrrc30

K4me3 loss in *Dnmt1*$^{-/-}$ MEFs (difference < -0.5194664 (for *Dnmt1*$^{-/-}$ - *Dnmt1*$^{+/+}$))					
Tceal1	Gm1993	D330045A20Rik	Zfp532	Stc1	4930519F16Rik
Sv2a	Flrt1	Zfhx4	Phyhd1	Cdo1	Itm2a
Creld1	Efemp1	Zfp503	Slx	Nr3c2	Sulf1
Gdpd2	Dlx5	Lpar4	Col4a5	Nap1l2	Gria3
Alx4	Pde9a	Ddx3y	Rxfp3	Gm5458	Cdh10
Ntf3	Gm14347	Mir301b	AW551984	Six2	Gm4297
Ldb2	Gm2863	Tmem176a	Olfr958	Tcfap2b	

Bibliography

Abdurashitov, M. A., V. A. Chernukhin, D. A. Gonchar and S. Degtyarev (2009). GlaI digestion of mouse gamma-satellite DNA: study of primary structure and ACGT sites methylation. *BMC Genomics.* 10: 322.

Agger, K., P. A. Cloos, J. Christensen, D. Pasini, S. Rose, J. Rappsilber, I. Issaeva, E. Canaani, A. E. Salcini and K. Helin (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature.* 449(7163): 731-734.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2002). Molecular Biology of the Cell. New York, Garland Science.

Alfano, G., I. Conte, T. Caramico, R. Avellino, B. Arno, M. T. Pizzo, N. Tanimoto, S. C. Beck, G. Huber, P. Dolle, M. W. Seeliger and S. Banfi (2011). Vax2 regulates retinoic acid distribution and cone opsin expression in the vertebrate eye. *Development.* 138(2): 261-271.

Allfrey, V. G., V. C. Littau and A. E. Mirsky (1963). On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. *Proc Natl Acad Sci U S A.* 49: 414-421.

Almer, A., H. Rudolph, A. Hinnen and W. Horz (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.* 5(10): 2689-2696.

Aran, D., G. Toperoff, M. Rosenberg and A. Hellman (2011). Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet.* 20(4): 670-680.

Aravin, A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor and G. J. Hannon (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell.* 31(6): 785-799.

Arita, K., M. Ariyoshi, H. Tochio, Y. Nakamura and M. Shirakawa (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature.* 455(7214): 818-821.

Atchison, L., A. Ghias, F. Wilkinson, N. Bonini and M. L. Atchison (2003). Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J.* 22(6): 1347-1358.

Augui, S., E. P. Nora and E. Heard (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet.* 12(6): 429-442.

Avraham, A., J. Sandbank, N. Yarom, A. Shalom, T. Karni, I. Pappo, A. Sella, A. Fich, S. Walfisch, L. Gheber and E. Evron (2010). A similar cell-specific pattern of HOXA methylation in normal and in cancer tissues. *Epigenetics.* 5(1): 41-46.

Avvakumov, G. V., J. R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C. H. Arrowsmith and S. Dhe-Paganon (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature.* 455(7214): 822-825.

Ball, M. P., J. B. Li, Y. Gao, J. H. Lee, E. M. LeProust, I. H. Park, B. Xie, G. Q. Daley and G. M. Church (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol.* 27(4): 361-368.

Banaszynski, L. A., C. D. Allis and P. W. Lewis (2010). Histone variants in metazoan development. *Dev Cell.* 19(5): 662-674.

Bantignies, F., V. Roure, I. Comet, B. Leblanc, B. Schuettengruber, J. Bonnet, V. Tixier, A. Mas and G. Cavalli (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell.* 144(2): 214-226.

Barr, H., A. Hermann, J. Berger, H. H. Tsai, K. Adie, A. Prokhortchouk, B. Hendrich and A. Bird (2007). Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol Cell Biol.* 27(10): 3750-3757.

Barreto, G., A. Schafer, J. Marhold, D. Stach, S. K. Swaminathan, V. Handa, G. Doderlein, N. Maltry, W. Wu, F. Lyko and C. Niehrs (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature.* 445(7128): 671-675.
Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). High-resolution profiling of histone methylations in the human genome. *Cell.* 129(4): 823-837.
Bartke, T., M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann and T. Kouzarides (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell.* 143(3): 470-484.
Beard, C., E. Li and R. Jaenisch (1995). Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev.* 9(19): 2325-2334.
Beisel, C. and R. Paro (2011). Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet.* 12(2): 123-135.
Bell, A. C. and G. Felsenfeld (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature.* 405(6785): 482-485.
Bender, W. and D. P. Fitzgerald (2002). Transcription activates repressed domains in the Drosophila bithorax complex. *Development.* 129(21): 4923-4930.
Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber and E. S. Lander (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell.* 125(2): 315-326.
Bertani, S., S. Sauer, E. Bolotin and F. Sauer (2011). The Noncoding RNA Mistral Activates Hoxa6 and Hoxa7 Expression and Stem Cell Differentiation by Recruiting MLL1 to Chromatin. *Mol Cell.* 43(6): 1040-1046.
Bhutani, N., J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel and H. M. Blau (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature.* 463(7284): 1042-1047.
Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16(1): 6-21.
Bird, A. (2007). Perceptions of epigenetics. *Nature.* 447(7143): 396-398.
Bird, A. (2011). The dinucleotide CG as a genomic signalling module. *J Mol Biol.* 409(1): 47-53.
Bird, A., M. Taggart, M. Frommer, O. J. Miller and D. Macleod (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell.* 40(1): 91-99.
Bird, A. P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* 11(3): 94-100.
Blackledge, N. P., J. C. Zhou, M. Y. Tolstorukov, A. M. Farcas, P. J. Park and R. J. Klose (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell.* 38(2): 179-190.
Blobel, G. A., S. Kadauke, E. Wang, A. W. Lau, J. Zuber, M. M. Chou and C. R. Vakoc (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol Cell.* 36(6): 970-983.
Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 19(2): 185-193.
Bonfils, C., N. Beaulieu, E. Chan, J. Cotton-Montpetit and A. R. MacLeod (2000). Characterization of the human DNA methyltransferase splice variant Dnmt1b. *J Biol Chem.* 275(15): 10754-10760.
Borgel, J., S. Guibert, Y. Li, H. Chiba, D. Schubeler, H. Sasaki, T. Forne and M. Weber (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet.* 42(12): 1093-1100.
Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan and S. E. Jacobsen (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science.* 317(5845): 1760-1764.
Bott, M., M. Brevet, B. S. Taylor, S. Shimizu, T. Ito, L. Wang, J. Creaney, R. A. Lake, M. F. Zakowski, B. Reva, C. Sander, R. Delsite, S. Powell, Q. Zhou, R. Shen, A. Olshen, V. Rusch and M. Ladanyi (2011). The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat Genet.* 43(7): 668-672.
Bourc'his, D. and T. H. Bestor (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature.* 431(7004): 96-99.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman and T. H. Bestor (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science.* 294(5551): 2536-2539.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, G. W. Bell, A. P. Otte, M. Vidal, D. K. Gifford, R. A. Young and R. Jaenisch (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature.* 441(7091): 349-353.

Boyes, J. and A. Bird (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* 11(1): 327-333.

Bracken, A. P., N. Dietrich, D. Pasini, K. H. Hansen and K. Helin (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20(9): 1123-1136.

Branciamore, S., Z. X. Chen, A. D. Riggs and S. N. Rodin (2010). CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. *Proc Natl Acad Sci U S A.* 107(35): 15485-15490.

Brenet, F., M. Moh, P. Funk, E. Feierstein, A. J. Viale, N. D. Socci and J. M. Scandura (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One.* 6(1): e14524.

Broske, A. M., L. Vockentanz, S. Kharazi, M. R. Huska, E. Mancini, M. Scheller, C. Kuhl, A. Enns, M. Prinz, R. Jaenisch, C. Nerlov, A. Leutz, M. A. Andrade-Navarro, S. E. Jacobsen and F. Rosenbauer (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet.* 41(11): 1207-1215.

Brunner, A. L., D. S. Johnson, S. W. Kim, A. Valouev, T. E. Reddy, N. F. Neff, E. Anton, C. Medina, L. Nguyen, E. Chiao, C. B. Oyolu, G. P. Schroth, D. M. Absher, J. C. Baker and R. M. Myers (2009). Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* 19(6): 1044-1056.

Buckle, R. S., J. D. Maman and J. Allan (1992). Site-directed mutagenesis studies on the binding of the globular domain of linker histone H5 to the nucleosome. *J Mol Biol.* 223(3): 651-659.

Burlingame, R. W., W. E. Love, B. C. Wang, R. Hamlin, H. X. Nguyen and E. N. Moudrianakis (1985). Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 A. *Science.* 228(4699): 546-553.

Butler, J. S., L. R. Palam, C. M. Tate, J. R. Sanford, R. C. Wek and D. G. Skalnik (2009). DNA Methyltransferase protein synthesis is reduced in CXXC finger protein 1-deficient embryonic stem cells. *DNA Cell Biol.* 28(5): 223-231.

Cabili, M. N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev and J. L. Rinn (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature.* 461(7261): 193-198.

Cao, R., Y. Tsukada and Y. Zhang (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell.* 20(6): 845-854.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones and Y. Zhang (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science.* 298(5595): 1039-1043.

Cao, R. and Y. Zhang (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell.* 15(1): 57-67.

Capotosti, F., J. J. Hsieh and W. Herr (2007). Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways. *Mol Cell Biol.* 27(20): 7063-7072.

Carlone, D. L., J. H. Lee, S. R. Young, E. Dobrota, J. S. Butler, J. Ruiz and D. G. Skalnik (2005). Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol Cell Biol.* 25(12): 4881-4891.

Casanova, M., T. Preissner, A. Cerase, R. Poot, D. Yamada, X. Li, R. Appanah, K. Bezstarosti, J. Demmers, H. Koseki and N. Brockdorff (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development.* 138(8): 1471-1482.

Casas-Delucchi, C. S., J. G. van Bemmel, S. Haase, H. D. Herce, D. Nowak, D. Meilinger, J. H. Stear, H. Leonhardt and M. C. Cardoso (2011). Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.*

Cavalli, G. and R. Paro (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell.* 93(4): 505-518.
Cavalli, G. and R. Paro (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science.* 286(5441): 955-958.
Chamberlain, S. J., D. Yee and T. Magnuson (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells.* 26(6): 1496-1505.
Chedin, F., M. R. Lieber and C. L. Hsieh (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A.* 99(26): 16916-16921.
Chen, L., A. M. MacMillan, W. Chang, K. Ezaz-Nikpay, W. S. Lane and G. L. Verdine (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry.* 30(46): 11018-11025.
Chen, R. Z., S. Akbarian, M. Tudor and R. Jaenisch (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genet.* 27(3): 327-331.
Chen, S. S., A. Raval, A. J. Johnson, E. Hertlein, T. H. Liu, V. X. Jin, M. H. Sherman, S. J. Liu, D. W. Dawson, K. E. Williams, M. Lanasa, S. Liyanarachchi, T. S. Lin, G. Marcucci, Y. Pekarsky, R. Davuluri, C. M. Croce, D. C. Guttridge, M. A. Teitell, J. C. Byrd and C. Plass (2009). Epigenetic changes during disease progression in a murine model of human chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A.* 106(32): 13433-13438.
Chen, T., N. Tsujimoto and E. Li (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol Cell Biol.* 24(20): 9048-9058.
Chen, T., Y. Ueda, S. Xie and E. Li (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J Biol Chem.* 277(41): 38746-38754.
Cheng, X. and R. M. Blumenthal (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure.* 16(3): 341-350.
Choi, D. K. and J. H. Do (2006). Normalization of microarray data: Single-labeled and dual-labeled arrays. *Mol Cells.* 22(3): 254-261.
Choi, J. K. and L. J. Howe (2009). Histone acetylation: truth of consequences? *Biochem Cell Biol.* 87(1): 139-150.
Chong, S., N. Vickaryous, A. Ashe, N. Zamudio, N. Youngson, S. Hemley, T. Stopka, A. Skoultchi, J. Matthews, H. S. Scott, D. de Kretser, M. O'Bryan, M. Blewitt and E. Whitelaw (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat Genet.* 39(5): 614-622.
Choy, J. S., S. Wei, J. Y. Lee, S. Tan, S. Chu and T. H. Lee (2010). DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc.* 132(6): 1782-1783.
Cierpicki, T., L. E. Risner, J. Grembecka, S. M. Lukasik, R. Popovic, M. Omonkowska, D. D. Shultis, N. J. Zeleznik-Le and J. H. Bushweller (2010). Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nat Struct Mol Biol.* 17(1): 62-68.
Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *J Am Stat Assoc.* 74(368): 829-836.
Cohen, N. M., E. Kenigsberg and A. Tanay (2011). Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell.* 145(5): 773-786.
Comet, I., B. Schuettengruber, T. Sexton and G. Cavalli (2011). A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc Natl Acad Sci U S A.* 108(6): 2294-2299.
Conerly, M. L., S. S. Teves, D. Diolaiti, M. Ulrich, R. N. Eisenman and S. Henikoff (2010). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res.* 20(10): 1383-1390.
Cooper, D. N., M. H. Taggart and A. P. Bird (1983). Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.* 11(3): 647-658.
Cortazar, D., C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird and P. Schar (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature.* 470(7334): 419-423.
Cortellino, S., J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T.

Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue and A. Bellacosa (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell.* 146(1): 67-79.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 107(50): 21931-21936.

Creyghton, M. P., S. Markoulaki, S. S. Levine, J. Hanna, M. A. Lodato, K. Sha, R. A. Young, R. Jaenisch and L. A. Boyer (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell.* 135(4): 649-661.

Cross, S. H., J. A. Charlton, X. Nan and A. P. Bird (1994). Purification of CpG islands using a methylated DNA binding column. *Nat Genet.* 6(3): 236-244.

Cross, S. H., R. R. Meehan, X. Nan and A. Bird (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet.* 16(3): 256-259.

Csankovszki, G., A. Nagy and R. Jaenisch (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol.* 153(4): 773-784.

Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof and V. Pirrotta (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell.* 111(2): 185-196.

Damelin, M. and T. H. Bestor (2007). Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol Cell Biol.* 27(11): 3891-3899.

Dawlaty, M. M., K. Ganz, B. E. Powell, Y. C. Hu, S. Markoulaki, A. W. Cheng, Q. Gao, J. Kim, S. W. Choi, D. C. Page and R. Jaenisch (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell.* 9(2): 166-175.

De La Fuente, R., C. Baumann, T. Fan, A. Schmidtmann, I. Dobrinski and K. Muegge (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nat Cell Biol.* 8(12): 1448-1454.

de Napoles, M., J. E. Mermoud, R. Wakao, Y. A. Tang, M. Endoh, R. Appanah, T. B. Nesterova, J. Silva, A. P. Otte, M. Vidal, H. Koseki and N. Brockdorff (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell.* 7(5): 663-676.

De Smet, C., O. De Backer, I. Faraoni, C. Lurquin, F. Brasseur and T. Boon (1996). The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl Acad Sci U S A.* 93(14): 7149-7153.

De Smet, C., C. Lurquin, B. Lethe, V. Martelange and T. Boon (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol.* 19(11): 7327-7335.

Deaton, A. M. and A. Bird (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25(10): 1010-1022.

Deaton, A. M., S. Webb, A. R. Kerr, R. S. Illingworth, J. Guy, R. Andrews and A. Bird (2011). Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* 21(7): 1074-1086.

Denis, H., M. N. Ndlovu and F. Fuks (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep.* 12(7): 647-656.

Dennis, K., T. Fan, T. Geiman, Q. Yan and K. Muegge (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* 15(22): 2940-2944.

Dhawan, S., S. Georgia, S. I. Tschen, G. Fan and A. Bhushan (2011). Pancreatic beta cell identity is maintained by DNA methylation-mediated repression of Arx. *Dev Cell.* 20(4): 419-429.

Dhayalan, A., A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin and A. Jeltsch (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem.* 285(34): 26114-26120.

Diehl, F., M. A. Brown, M. J. van Amerongen, T. Novoyatleva, A. Wietelmann, J. Harriss, F. Ferrazzi, T. Bottger, R. P. Harvey, P. W. Tucker and F. B. Engel (2010). Cardiac deletion of Smyd2 is dispensable for mouse heart development. *PLoS One.* 5(3): e9748.

Ding, F. and J. R. Chaillet (2002). In vivo stabilization of the Dnmt1 (cytosine-5)- methyltransferase protein. *Proc Natl Acad Sci U S A.* 99(23): 14861-14866.

Dinger, M. E., P. P. Amaral, T. R. Mercer, K. C. Pang, S. J. Bruce, B. B. Gardiner, M. E. Askarian-Amiri, K. Ru, G. Solda, C. Simons, S. M. Sunkin, M. L. Crowe, S. M. Grimmond, A. C. Perkins and J. S. Mattick (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18(9): 1433-1445.

Di-Poi, N., J. I. Montoya-Burgos, H. Miller, O. Pourquie, M. C. Milinkovitch and D. Duboule (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature.* 464(7285): 99-103.

Doi, A., I. H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley and A. P. Feinberg (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet.* 41(12): 1350-1353.

Dong, K. B., I. A. Maksakova, F. Mohn, D. Leung, R. Appanah, S. Lee, H. W. Yang, L. L. Lam, D. L. Mager, D. Schubeler, M. Tachibana, Y. Shinkai and M. C. Lorincz (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* 27(20): 2691-2701.

Down, T. A., V. K. Rakyan, D. J. Turner, P. Flicek, H. Li, E. Kulesha, S. Graf, N. Johnson, J. Herrero, E. M. Tomazou, N. P. Thorne, L. Backdahl, M. Herberth, K. L. Howe, D. K. Jackson, M. M. Miretti, J. C. Marioni, E. Birney, T. J. Hubbard, R. Durbin, S. Tavare and S. Beck (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol.* 26(7): 779-785.

Dunican, D. S., A. Ruzov, J. A. Hackett and R. R. Meehan (2008). xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function. *Development.* 135(7): 1295-1302.

Easwaran, H. P., L. Schermelleh, H. Leonhardt and M. C. Cardoso (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* 5(12): 1181-1186.

Eckhardt, F., J. Lewin, R. Cortese, V. K. Rakyan, J. Attwood, M. Burger, J. Burton, T. V. Cox, R. Davies, T. A. Down, C. Haefliger, R. Horton, K. Howe, D. K. Jackson, J. Kunde, C. Koenig, J. Liddle, D. Niblett, T. Otto, R. Pettett, S. Seemann, C. Thompson, T. West, J. Rogers, A. Olek, K. Berlin and S. Beck (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* 38(12): 1378-1385.

Elderkin, S., G. N. Maertens, M. Endoh, D. L. Mallery, N. Morrice, H. Koseki, G. Peters, N. Brockdorff and K. Hiom (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Mol Cell.* 28(1): 107-120.

Epsztejn-Litman, S., N. Feldman, M. Abu-Remaileh, Y. Shufaro, A. Gerson, J. Ueda, R. Deplus, F. Fuks, Y. Shinkai, H. Cedar and Y. Bergman (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol.* 15(11): 1176-1183.

Eskeland, R., M. Leeb, G. R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A. I. Skoultchi, A. Wutz and W. A. Bickmore (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell.* 38(3): 452-464.

Espada, J., H. Peinado, L. Lopez-Serra, F. Setien, P. Lopez-Serra, A. Portela, J. Renart, E. Carrasco, M. Calvo, A. Juarranz, A. Cano and M. Esteller (2011). Regulation of SNAIL1 and E-cadherin function by DNMT1 in a DNA methylation-independent context. *Nucleic Acids Res.*

Esteve, P. O., Y. Chang, M. Samaranayake, A. K. Upadhyay, J. R. Horton, G. R. Feehery, X. Cheng and S. Pradhan (2011). A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nat Struct Mol Biol.* 18(1): 42-48.

Esteve, P. O., H. G. Chin, A. Smallwood, G. R. Feehery, O. Gangisetty, A. R. Karpf, M. F. Carey and S. Pradhan (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* 20(22): 3089-3103.

Faast, R., V. Thonglairoam, T. C. Schulz, J. Beall, J. R. Wells, H. Taylor, K. Matthaei, P. D. Rathjen, D. J. Tremethick and I. Lyons (2001). Histone variant H2A.Z is required for early mammalian development. *Curr Biol.* 11(15): 1183-1187.

Fan, Y., T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock and A. I. Skoultchi (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell.* 123(7): 1199-1212.

Farthing, C. R., G. Ficz, R. K. Ng, C. F. Chan, S. Andrews, W. Dean, M. Hemberger and W. Reik (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* 4(6): e1000116.

Feldman, N., A. Gerson, J. Fang, E. Li, Y. Zhang, Y. Shinkai, H. Cedar and Y. Bergman (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol.* 8(2): 188-194.

Felle, M., H. Hoffmeister, J. Rothammer, A. Fuchs, J. H. Exler and G. Langst (2011b). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res.*

Felle, M., S. Joppien, A. Nemeth, S. Diermeier, V. Thalhammer, T. Dobner, E. Kremmer, R. Kappler and G. Langst (2011a). The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res.*

Feng, S., S. J. Cokus, X. Zhang, P. Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini and S. E. Jacobsen (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A.* 107(19): 8689-8694.

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet.* 12(8): 565-575.

Ferraiuolo, M. A., M. Rousseau, C. Miyamoto, S. Shenker, X. Q. Wang, M. Nadler, M. Blanchette and J. Dostie (2010). The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* 38(21): 7472-7484.

Ficz, G., M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews and W. Reik (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature.* 473(7347): 398-402.

Filion, G. J., S. Zhenilo, S. Salozhin, D. Yamada, E. Prokhortchouk and P. A. Defossez (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol.* 26(1): 169-181.

Fouse, S. D., Y. Shen, M. Pellegrini, S. Cole, A. Meissner, L. Van Neste, R. Jaenisch and G. Fan (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell.* 2(2): 160-169.

Francis, N. J., N. E. Follmer, M. D. Simon, G. Aghia and J. D. Butler (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell.* 137(1): 110-122.

Frauer, C., T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes and H. Leonhardt (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One.* 6(6): e21306.

Fujimura, Y., K. Isono, M. Vidal, M. Endoh, H. Kajita, Y. Mizutani-Koseki, Y. Takihara, M. van Lohuizen, A. Otte, T. Jenuwein, J. Deschamps and H. Koseki (2006). Distinct roles of Polycomb group gene products in transcriptionally repressed and active domains of Hoxb8. *Development.* 133(12): 2371-2381.

Fujita, N., S. Watanabe, T. Ichimura, S. Tsuruzoe, Y. Shinkai, M. Tachibana, T. Chiba and M. Nakao (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J Biol Chem.* 278(26): 24132-24138.

Galli, A., D. Robay, M. Osterwalder, X. Bao, J. D. Benazet, M. Tariq, R. Paro, S. Mackem and R. Zeller (2010). Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. *PLoS Genet.* 6(4): e1000901.

Gao, J., J. Wang, Y. Wang, W. Dai and L. Lu (2011). Regulation of Pax6 by CTCF during induction of mouse ES cell differentiation. *PLoS One.* 6(6): e20954.

Garrick, D., M. De Gobbi, V. Samara, M. Rugless, M. Holland, H. Ayyub, K. Lower, J. Sloane-Stanley, N. Gray, C. Koch, I. Dunham and D. R. Higgs (2008). The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. *Blood.* 112(9): 3889-3899.

Ge, Y. Z., M. T. Pu, H. Gowher, H. P. Wu, J. P. Ding, A. Jeltsch and G. L. Xu (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem.* 279(24): 25447-25454.

Geahigan, K. B., G. A. Meints, M. E. Hatcher, J. Orban and G. P. Drobny (2000). The dynamic impact of CpG methylation in DNA. *Biochemistry.* 39(16): 4939-4946.

Gehani, S. S., S. Agrawal-Singh, N. Dietrich, N. S. Christophersen, K. Helin and K. Hansen (2010). Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. *Mol Cell.* 39(6): 886-900.

Geiman, T. M., S. K. Durum and K. Muegge (1998). Characterization of gene expression, genomic structure, and chromosomal localization of Hells (Lsh). *Genomics.* 54(3): 477-483.

Geiman, T. M., U. T. Sankpal, A. K. Robertson, Y. Zhao and K. D. Robertson (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem Biophys Res Commun.* 318(2): 544-555.

Geiman, T. M., L. Tessarollo, M. R. Anver, J. B. Kopp, J. M. Ward and K. Muegge (2001). Lsh, a SNF2 family member, is required for normal murine development. *Biochim Biophys Acta.* 1526(2): 211-220.

Gibbons, R. J., T. L. McDowell, S. Raman, D. M. O'Rourke, D. Garrick, H. Ayyub and D. R. Higgs (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet.* 24(4): 368-371.

Gilbert, N., I. Thomson, S. Boyle, J. Allan, B. Ramsahoye and W. A. Bickmore (2007). DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol.* 177(3): 401-411.

Goll, M. G. and T. H. Bestor (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem.* 74: 481-514.

Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen and T. H. Bestor (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science.* 311(5759): 395-398.

Goto, Y. and H. Kimura (2009). Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. *Nucleic Acids Res.* 37(22): 7416-7428.

Gou, D., M. Rubalcava, S. Sauer, F. Mora-Bermudez, H. Erdjument-Bromage, P. Tempst, E. Kremmer and F. Sauer (2010). SETDB1 is involved in postembryonic DNA methylation and gene silencing in Drosophila. *PLoS One.* 5(5): e10581.

Gruss, C. and J. M. Sogo (1992). Chromatin replication. *Bioessays.* 14(1): 1-8.

Gu, T. P., F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li and G. L. Xu (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature.*

Guo, J. U., Y. Su, C. Zhong, G. L. Ming and H. Song (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell.* 145(3): 423-434.

Guy, J., B. Hendrich, M. Holmes, J. E. Martin and A. Bird (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet.* 27(3): 322-326.

Ha, K., G. E. Lee, S. S. Palii, K. D. Brown, Y. Takeda, K. Liu, K. N. Bhalla and K. D. Robertson (2011). Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum Mol Genet.* 20(1): 126-140.

Hackett, J. A. (2010). Epigenetic Regulation of Germline-Specific Genes. *PhD thesis.*

Hajkova, P., K. Ancelin, T. Waldmann, N. Lacoste, U. C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider and M. A. Surani (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature.* 452(7189): 877-881.

Hajkova, P., S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson and M. A. Surani (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science.* 329(5987): 78-82.

Handa, V. and A. Jeltsch (2005). Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J Mol Biol.* 348(5): 1103-1112.

Hansen, K. D., W. Timp, H. C. Bravo, S. Sabunciyan, B. Langmead, O. G. McDonald, B. Wen, H. Wu, Y. Liu, D. Diep, E. Briem, K. Zhang, R. A. Irizarry and A. P. Feinberg (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet.* 43(8): 768-775.

Hansen, K. H., A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani, A. Monrad, J. Rappsilber, M. Lerdrup and K. Helin (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol.* 10(11): 1291-1300.

Harikrishnan, K. N., M. Z. Chow, E. K. Baker, S. Pal, S. Bassal, D. Brasacchio, L. Wang, J. M. Craig, P. L. Jones, S. Sif and A. El-Osta (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet.* 37(3): 254-264.

Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse and S. M. Tilghman (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature.* 405(6785): 486-489.

Hashimoto, H., J. R. Horton, X. Zhang and X. Cheng (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics.* 4(1): 8-14.

Hata, K., M. Okano, H. Lei and E. Li (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development.* 129(8): 1983-1993.

Havecker, E. R., X. Gao and D. F. Voytas (2004). The diversity of LTR retrotransposons. *Genome Biol.* 5(6): 225.

Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker and B. Ren (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell.* 6(5): 479-491.

He, H., X. Hua and J. Yan (2011b). Epigenetic regulations in hematopoietic Hox code. *Oncogene.* 30(4): 379-388.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He and G. L. Xu (2011a). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science.* 333(6047): 1303-1307.

Heard, E., C. Rougeulle, D. Arnaud, P. Avner, C. D. Allis and D. L. Spector (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell.* 107(6): 727-738.

Hekimoglu, B. and L. Ringrose (2009). Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol.* 6(2): 129-137.

Hellman, A. and A. Chess (2007). Gene body-specific methylation on the active X chromosome. *Science.* 315(5815): 1141-1143.

Hendrich, B. and A. Bird (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol.* 18(11): 6538-6547.

Hendrich, B., J. Guy, B. Ramsahoye, V. A. Wilson and A. Bird (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* 15(6): 710-723.

Hermann, A., S. Schmitt and A. Jeltsch (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J Biol Chem.* 278(34): 31717-31721.

Hernandez-Munoz, I., P. Taghavi, C. Kuijl, J. Neefjes and M. van Lohuizen (2005). Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol Cell Biol.* 25(24): 11047-11058.

Ho, K. L., I. W. McNae, L. Schmiedeberg, R. J. Klose, A. P. Bird and M. D. Walkinshaw (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell.* 29(4): 525-531.

Hochedlinger, K. and K. Plath (2009). Epigenetic reprogramming and induced pluripotency. *Development.* 136(4): 509-523.

Hodges, E., A. D. Smith, J. Kendall, Z. Xuan, K. Ravi, M. Rooks, M. Q. Zhang, K. Ye, A. Bhattacharjee, L. Brizuela, W. R. McCombie, M. Wigler, G. J. Hannon and J. B. Hicks (2009). High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res.* 19(9): 1593-1605.

Hodgson, J. W. and H. W. Brock (2011). Are polycomb group bodies gene silencing factories? *Cell.* 144(2): 170-171.

Hogga, I. and F. Karch (2002). Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of Polycomb-mediated silencing. *Development.* 129(21): 4915-4922.

Holliday, R. and J. E. Pugh (1975). DNA modification mechanisms and gene activity during development. *Science.* 187(4173): 226-232.

Hsieh, J. J., E. H. Cheng and S. J. Korsmeyer (2003). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell.* 115(3): 293-303.

Hsu, D. W., M. J. Lin, T. L. Lee, S. C. Wen, X. Chen and C. K. Shen (1999). Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. *Proc Natl Acad Sci U S A.* 96(17): 9751-9756.

Hu, K., X. Nan, A. Bird and W. Wang (2006). Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex. *Nat Genet.* 38(9): 962-964; author reply 964-967.

Hu, M., X. J. Sun, Y. L. Zhang, Y. Kuang, C. Q. Hu, W. L. Wu, S. H. Shen, T. T. Du, H. Li, F. He, H. S. Xiao, Z. G. Wang, T. X. Liu, H. Lu, Q. H. Huang, S. J. Chen and Z. Chen (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc Natl Acad Sci U S A.* 107(7): 2956-2961.

Huang, J., T. Fan, Q. Yan, H. Zhu, S. Fox, H. J. Issaq, L. Best, L. Gangi, D. Munroe and K. Muegge (2004). Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res.* 32(17): 5019-5028.

Huang, R. C. and J. Bonner (1962). Histone, a suppressor of chromosomal RNA synthesis. *Proc Natl Acad Sci U S A.* 48: 1216-1222.

Illingworth, R., A. Kerr, D. Desousa, H. Jorgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford and A. Bird (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 6(1): e22.

Illingworth, R. S. and A. P. Bird (2009). CpG islands--'a rough guide'. *FEBS Lett.* 583(11): 1713-1720.

Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews and A. P. Bird (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6(9).

Inoue, A. and Y. Zhang (2011). Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science.*

Irizarry, R. A., C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan and A. P. Feinberg (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet.* 41(2): 178-186.

Isono, K., Y. Fujimura, J. Shinga, M. Yamaki, O. W. J, Y. Takihara, Y. Murahashi, Y. Takada, Y. Mizutani-Koseki and H. Koseki (2005). Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. *Mol Cell Biol.* 25(15): 6694-6706.

Issaeva, I., Y. Zonis, T. Rozovskaia, K. Orlovsky, C. M. Croce, T. Nakamura, A. Mazo, L. Eisenbach and E. Canaani (2007). Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol.* 27(5): 1889-1903.

Ito, S., A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers and Y. Zhang (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature.* 466(7310): 1129-1133.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science.*

Jackson, M., A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell and B. Ramsahoye (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol.* 24(20): 8862-8871.

Jackson-Grusby, L., C. Beard, R. Possemato, M. Tudor, D. Fambrough, G. Csankovszki, J. Dausman, P. Lee, C. Wilson, E. Lander and R. Jaenisch (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet.* 27(1): 31-39.

Jenuwein, T. and C. D. Allis (2001). Translating the histone code. *Science.* 293(5532): 1074-1080.

Jeong, S., G. Liang, S. Sharma, J. C. Lin, S. H. Choi, H. Han, C. B. Yoo, G. Egger, A. S. Yang and P. A. Jones (2009). Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol Cell Biol.* 29(19): 5366-5376.

Ji, H., L. I. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M. J. Aryee, R. A. Irizarry, K. Kim, D. J. Rossi, M. A. Inlay, T. Serwold, H. Karsunky, L. Ho, G. Q. Daley, I. L. Weissman and A. P. Feinberg (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature.* 467(7313): 338-342.

Jia, D., R. Z. Jurkowska, X. Zhang, A. Jeltsch and X. Cheng (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature.* 449(7159): 248-251.

Jin, B., Q. Tao, J. Peng, H. M. Soo, W. Wu, J. Ying, C. R. Fields, A. L. Delmas, X. Liu, J. Qiu and K. D. Robertson (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum Mol Genet.* 17(5): 690-709.

Jin, B., B. Yao, J. L. Li, C. R. Fields, A. L. Delmas, C. Liu and K. D. Robertson (2009). DNMT1 and DNMT3B modulate distinct polycomb-mediated histone modifications in colon cancer. *Cancer Res.* 69(18): 7412-7421.

Jin, S. G., C. Guo and G. P. Pfeifer (2008). GADD45A does not promote DNA demethylation. *PLoS Genet.* 4(3): e1000013.

Jones, P. A. and G. Liang (2009). Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet.* 10(11): 805-811.

Jones, P. L., G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis and A. P. Wolffe (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet.* 19(2): 187-191.

Jorgensen, H. F., I. Ben-Porath and A. P. Bird (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol Cell Biol.* 24(8): 3387-3395.

Kafri, T., M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar and A. Razin (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* 6(5): 705-714.

Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li and H. Sasaki (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature.* 429(6994): 900-903.

Kanhere, A., K. Viiri, C. C. Araujo, J. Rasaiyaah, R. D. Bouwman, W. A. Whyte, C. F. Pereira, E. Brookes, K. Walker, G. W. Bell, A. Pombo, A. G. Fisher, R. A. Young and R. G. Jenner (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell.* 38(5): 675-688.

Kareta, M. S., Z. M. Botello, J. J. Ennis, C. Chou and F. Chedin (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J Biol Chem.* 281(36): 25893-25902.

Karimi, M. M., P. Goyal, I. A. Maksakova, M. Bilenky, D. Leung, J. X. Tang, Y. Shinkai, D. L. Mager, S. Jones, M. Hirst and M. C. Lorincz (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell.* 8(6): 676-687.

Kashiwagi, K., K. Nimura, K. Ura and Y. Kaneda (2011). DNA methyltransferase 3b preferentially associates with condensed chromatin. *Nucleic Acids Res.* 39(3): 874-888.

Khulan, B., R. F. Thompson, K. Ye, M. J. Fazzari, M. Suzuki, E. Stasiek, M. E. Figueroa, J. L. Glass, Q. Chen, C. Montagna, E. Hatchwell, R. R. Selzer, T. A. Richmond, R. D. Green, A. Melnick and J. M. Greally (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* 16(8): 1046-1055.

Kim, G. D., J. Ni, N. Kelesoglu, R. J. Roberts and S. Pradhan (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J.* 21(15): 4183-4195.

Kim, H., K. Kang and J. Kim (2009b). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* 37(9): 2940-2950.

Kim, M. S., T. Kondo, I. Takada, M. Y. Youn, Y. Yamamoto, S. Takahashi, T. Matsumoto, S. Fujiyama, Y. Shirode, I. Yamaoka, H. Kitagawa, K. Takeyama, H. Shibuya, F. Ohtake and S. Kato (2009a). DNA demethylation in hormone-induced transcriptional derepression. *Nature.* 461(7266): 1007-1012.

Kim, S. W., J. I. Park, C. M. Spring, A. K. Sater, H. Ji, A. A. Otchere, J. M. Daniel and P. D. McCrea (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat Cell Biol.* 6(12): 1212-1220.

Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov, R. D. Green, M. Q. Zhang, V. V. Lobanenkov and B. Ren (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell.* 128(6): 1231-1245.

Kim, Y. J., K. R. Cecchini and T. H. Kim (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc Natl Acad Sci U S A.* 108(18): 7391-7396.

Klein, C. J., M. V. Botuyan, Y. Wu, C. J. Ward, G. A. Nicholson, S. Hammans, K. Hojo, H. Yamanishi, A. R. Karpf, D. C. Wallace, M. Simon, C. Lander, L. A. Boardman, J. M. Cunningham, G. E. Smith, W. J. Litchy, B. Boes, E. J. Atkinson, S. Middha, B. D. PJ, J. E. Parisi, G. Mer, D. I. Smith and P. J. Dyck (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat Genet.* 43(6): 595-600.

Klose, R. J. and A. P. Bird (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem.* 279(45): 46490-46496.

Klose, R. J. and A. P. Bird (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci.* 31(2): 89-97.

Klose, R. J., S. A. Sarraf, L. Schmiedeberg, S. M. McDermott, I. Stancheva and A. P. Bird (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell.* 19(5): 667-678.

Ko, M., H. S. Bandukwala, J. An, E. D. Lamperti, E. C. Thompson, R. Hastie, A. Tsangaratou, K. Rajewsky, S. B. Koralov and A. Rao (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A.* 108(35): 14566-14571.

Ko, M., Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski and A. Rao (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature.* 468(7325): 839-843.

Koh, K. P., A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley and A. Rao (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell.* 8(2): 200-213.

Kokura, K., S. C. Kaul, R. Wadhwa, T. Nomura, M. M. Khan, T. Shinagawa, T. Yasukawa, C. Colmenares and S. Ishii (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. *J Biol Chem.* 276(36): 34115-34121.

Komashko, V. M. and P. J. Farnham (2010). 5-azacytidine treatment reorganizes genomic histone modification patterns. *Epigenetics.* 5(3).

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell.* 128(4): 693-705.

Kriaucionis, S. and N. Heintz (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science.* 324(5929): 929-930.

Ku, M., R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh, T. S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A. S. Chi, M. Adli, S. Kasif, L. M. Ptaszek, C. A. Cowan, E. S. Lander, H. Koseki and B. E. Bernstein (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4(10): e1000242.

Kumaki, Y., M. Oda and M. Okano (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* 36(Web Server issue): W170-175.

Kunert, N., J. Marhold, J. Stanke, D. Stach and F. Lyko (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development.* 130(21): 5083-5090.

Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki and T. Nakano (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 22(7): 908-917.

Kurdistani, S. K., S. Tavazoie and M. Grunstein (2004). Mapping global histone acetylation patterns to gene expression. *Cell.* 117(6): 721-733.

Laget, S., M. Joulie, F. Le Masson, N. Sasai, E. Christians, S. Pradhan, R. J. Roberts and P. A. Defossez (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One.* 5(8): e11982.

Lai, A. Y., M. Fatemi, A. Dhasarathy, C. Malone, S. E. Sobol, C. Geigerman, D. L. Jaye, D. Mav, R. Shah, L. Li and P. A. Wade (2010). DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med.* 207(9): 1939-1950.

Laird, P. W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet.* 11(3): 191-203.

Lan, F., P. E. Bayliss, J. L. Rinn, J. R. Whetstine, J. K. Wang, S. Chen, S. Iwase, R. Alpatov, I. Issaeva, E. Canaani, T. M. Roberts, H. Y. Chang and Y. Shi (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature*. 449(7163): 689-694.

Lande-Diner, L., J. Zhang, I. Ben-Porath, N. Amariglio, I. Keshet, M. Hecht, V. Azuara, A. G. Fisher, G. Rechavi and H. Cedar (2007). Role of DNA methylation in stable gene repression. *J Biol Chem.* 282(16): 12194-12200.

Lau, P. N. and P. Cheung (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proc Natl Acad Sci U S A.* 108(7): 2801-2806.

Laurent, L., E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring and C. L. Wei (2010). Dynamic changes in the human methylome during differentiation. *Genome Res.* 20(3): 320-331.

Lee, J. H., K. S. Voo and D. G. Skalnik (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem.* 276(48): 44669-44676.

Lee, T. I., R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine, R. M. Kumar, B. Chevalier, S. E. Johnstone, M. F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H. L. Murray, J. P. Zucker, B. Yuan, G. W. Bell, E. Herbolsheimer, N. M. Hannett, K. Sun, D. T. Odom, A. P. Otte, T. L. Volkert, D. P. Bartel, D. A. Melton, D. K. Gifford, R. Jaenisch and R. A. Young (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell.* 125(2): 301-313.

Leeb, M., D. Pasini, M. Novatchkova, M. Jaritz, K. Helin and A. Wutz (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* 24(3): 265-276.

Lefevre, P., J. Witham, C. E. Lacroix, P. N. Cockerill and C. Bonifer (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol Cell.* 32(1): 129-139.

Lehnertz, B., Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein and A. H. Peters (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol.* 13(14): 1192-1200.

Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch and E. Li (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development.* 122(10): 3195-3205.

Leonhardt, H., A. W. Page, H. U. Weier and T. H. Bestor (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell.* 71(5): 865-873.

Lewis, J. D., R. R. Meehan, W. J. Henzel, I. Maurer-Fogy, P. Jeppesen, F. Klein and A. Bird (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell.* 69(6): 905-914.

Ley, T. J., L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch, C. C. Harris, C. F. Lichti, R. R. Townsend, R. S. Fulton, D. J. Dooling, D. C. Koboldt, H. Schmidt, Q. Zhang, J. R. Osborne, L. Lin, M. O'Laughlin, J. F. McMichael, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, V. J. Magrini, T. L. Vickery, J. Hundal, L. L. Cook, J. J. Conyers, G. W. Swift, J. P. Reed, P. A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M. A. Watson, S. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M. H. Tomasson, D. C. Link, T. A. Graubert, J. F. DiPersio, E. R. Mardis and R. K. Wilson (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* 363(25): 2424-2433.

Li, B., M. Carey and J. L. Workman (2007). The role of chromatin during transcription. *Cell.* 128(4): 707-719.

Li, E., C. Beard and R. Jaenisch (1993). Role for DNA methylation in genomic imprinting. *Nature.* 366(6453): 362-365.

Li, E., T. H. Bestor and R. Jaenisch (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 69(6): 915-926.

Li, F., R. Martienssen and W. Z. Cande (2011a). Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. *Nature.* 475(7355): 244-248.

Li, G., R. Margueron, M. Ku, P. Chambon, B. E. Bernstein and D. Reinberg (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24(4): 368-380.

Li, H. B., M. Muller, I. A. Bahechar, O. Kyrchanova, K. Ohno, P. Georgiev and V. Pirrotta (2011b). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. *Mol Cell Biol.* 31(4): 616-625.

Li, T., J. F. Hu, X. Qiu, J. Ling, H. Chen, S. Wang, A. Hou, T. H. Vu and A. R. Hoffman (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol.* 28(20): 6473-6482.

Li, Y., J. Zhu, G. Tian, N. Li, Q. Li, M. Ye, H. Zheng, J. Yu, H. Wu, J. Sun, H. Zhang, Q. Chen, R. Luo, M. Chen, Y. He, X. Jin, Q. Zhang, C. Yu, G. Zhou, Y. Huang, H. Cao, X. Zhou, S. Guo, X. Hu, X. Li, K. Kristiansen, L. Bolund, J. Xu, W. Wang, H. Yang, J. Wang, R. Li, S. Beck and X. Zhang (2010). The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* 8(11): e1000533.

Li, Z., X. Cai, C. Cai, J. Wang, W. Zhang, B. E. Petersen, F. C. Yang and M. Xu (2011c). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood.*

Lindroth, A. M., Y. J. Park, C. M. McLean, G. A. Dokshin, J. M. Persson, H. Herman, D. Pasini, X. Miro, M. E. Donohoe, J. T. Lee, K. Helin and P. D. Soloway (2008). Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS Genet.* 4(8): e1000145.

Ling, J. Q., A. Hou and A. R. Hoffman (2011). Long-range DNA interactions are specifically altered by locked nucleic acid-targeting of a CTCF binding site. *Biochim Biophys Acta.* 1809(1): 24-33.

Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar and J. R. Ecker (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell.* 133(3): 523-536.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 462(7271): 315-322.

Lorch, Y., M. Zhang and R. D. Kornberg (1999). Histone octamer transfer by a chromatin-remodeling complex. *Cell.* 96(3): 389-392.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature.* 389(6648): 251-260.

Lynch, M. D., A. J. Smith, M. De Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon, R. J. Gibbons, D. Garrick and D. R. Higgs (2011). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.*

Maatouk, D. M., L. D. Kellam, M. R. Mann, H. Lei, E. Li, M. S. Bartolomei and J. L. Resnick (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development.* 133(17): 3411-3418.

Macleod, D., V. H. Clark and A. Bird (1999). Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nat Genet.* 23(2): 139-140.

Margueron, R., N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. Drury, 3rd, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg and S. J. Gamblin (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature.* 461(7265): 762-767.

Margueron, R. and D. Reinberg (2011). The Polycomb complex PRC2 and its mark in life. *Nature.* 469(7330): 343-349.

Martin Caballero, I., J. Hansen, D. Leaford, S. Pollard and B. D. Hendrich (2009). The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS One.* 4(1): e4315.

Martin, C. C., L. Laforest, M. A. Akimenko and M. Ekker (1999). A role for DNA methylation in gastrulation and somite patterning. *Dev Biol.* 206(2): 189-205.

Matarazzo, M. R., S. Boyle, M. D'Esposito and W. A. Bickmore (2007). Chromosome territory reorganization in a human disease with altered DNA methylation. *Proc Natl Acad Sci U S A.* 104(42): 16546-16551.

Mathieu, O., J. Reinders, M. Caikovski, C. Smathajitt and J. Paszkowski (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell.* 130(5): 851-862.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang and J. F. Costello (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 466(7303): 253-257.

McBryant, S. J., X. Lu and J. C. Hansen (2010). Multifunctionality of the linker histones: an emerging role for protein-protein interactions. *Cell Res.* 20(5): 519-528.

Meehan, R. R., J. D. Lewis and A. P. Bird (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* 20(19): 5085-5092.

Meehan, R. R., J. D. Lewis, S. McKay, E. L. Kleiner and A. P. Bird (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell.* 58(3): 499-507.

Meilinger, D., K. Fellinger, S. Bultmann, U. Rothbauer, I. M. Bonapace, W. E. Klinkert, F. Spada and H. Leonhardt (2009). Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.* 10(11): 1259-1264.

Meints, G. A. and G. P. Drobny (2001). Dynamic impact of methylation at the M. Hhai target site: a solid-state deuterium NMR study. *Biochemistry.* 40(41): 12436-12443.

Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch and E. S. Lander (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature.* 454(7205): 766-770.

Mendenhall, E. M., R. P. Koche, T. Truong, V. W. Zhou, B. Issac, A. S. Chi, M. Ku and B. E. Bernstein (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 6(12): e1001244.

Meselson, M. and F. W. Stahl (1958). The Replication of DNA in Escherichia Coli. *Proc Natl Acad Sci U S A.* 44(7): 671-682.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander and B. E. Bernstein (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 448(7153): 553-560.

Miranda, T. B. and P. A. Jones (2007). DNA methylation: the nuts and bolts of repression. *J Cell Physiol.* 213(2): 384-390.

Moarefi, A. H. and F. Chedin (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *J Mol Biol.* 409(5): 758-772.

Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel and D. Schubeler (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell.* 30(6): 755-766.

Monk, M., M. Boubelik and S. Lehnert (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development.* 99(3): 371-382.

Moran-Crusio, K., L. Reavie, A. Shih, O. Abdel-Wahab, D. Ndiaye-Lobry, C. Lobry, M. E. Figueroa, A. Vasanthakumar, J. Patel, X. Zhao, F. Perna, S. Pandey, J. Madzo, C. Song, Q. Dai, C. He, S. Ibrahim, M. Beran, J. Zavadil, S. D. Nimer, A. Melnick, L. A. Godley, I. Aifantis and R. L. Levine (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell.* 20(1): 11-24.

Muotri, A. R., M. C. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima and F. H. Gage (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature.* 468(7322): 443-446.

Myant, K., A. Termanis, A. Y. Sundaram, T. Boe, C. Li, C. Merusi, J. Burrage, J. I. de Las Heras and I. Stancheva (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.* 21(1): 83-94.

Nady, N., A. Lemak, J. R. Walker, G. V. Avvakumov, M. S. Kareta, M. Achour, S. Xue, S. Duan, A. Allali-Hassani, X. Zuo, Y. X. Wang, C. Bronner, F. Chedin, C. H. Arrowsmith and S. Dhe-Paganon (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J Biol Chem.* 286(27): 24300-24311.

Nagano, T., J. A. Mitchell, L. A. Sanz, F. M. Pauler, A. C. Ferguson-Smith, R. Feil and P. Fraser (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science.* 322(5908): 1717-1720.

Nan, X., R. R. Meehan and A. Bird (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* 21(21): 4886-4892.

Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman and A. Bird (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature.* 393(6683): 386-389.

Nathan, D. and D. M. Crothers (2002). Bending and flexibility of methylated and unmethylated EcoRI DNA. *J Mol Biol.* 316(1): 7-17.

Ndlovu, N., H. Denis and F. Fuks (2011). Exposing the DNA methylome iceberg. *Trends Biochem Sci.* 36(7): 381-387.

Negishi, M., A. Saraya, S. Miyagi, K. Nagao, Y. Inagaki, M. Nishikawa, S. Tajima, H. Koseki, H. Tsuda, Y. Takasaki, H. Nakauchi and A. Iwama (2007). Bmi1 cooperates with Dnmt1-associated protein 1 in gene silencing. *Biochem Biophys Res Commun.* 353(4): 992-998.

Nestor, C., A. Ruzov, R. Meehan and D. Dunican (2010). Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques.* 48(4): 317-319.

Nestor, C. E., R. Ottaviano, J. Reddington, D. Sproul, D. Reinhardt, D. Dunican, E. Katz, J. M. Dixon, D. J. Harrison and R. Meehan (2011). Tissue-type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.*

Ng, H. H., Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg and A. Bird (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet.* 23(1): 58-61.

Ng, R. K., W. Dean, C. Dawson, D. Lucifero, Z. Madeja, W. Reik and M. Hemberger (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol.* 10(11): 1280-1290.

Noordermeer, D., M. Leleu, E. Splinter, J. Rougemont, W. De Laat and D. Duboule (2011). The dynamic architecture of Hox gene clusters. *Science.* 334(6053): 222-225.

O'Carroll, D., S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani and T. Jenuwein (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol.* 21(13): 4330-4336.

Oda, M., J. L. Glass, R. F. Thompson, Y. Mo, E. N. Olivier, M. E. Figueroa, R. R. Selzer, T. A. Richmond, X. Zhang, L. Dannenberg, R. D. Green, A. Melnick, E. Hatchwell, E. E. Bouhassira, A. Verma, M. Suzuki and J. M. Greally (2009). High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* 37(12): 3829-3839.

O'Gara, M., S. Klimasauskas, R. J. Roberts and X. Cheng (1996). Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes. *J Mol Biol.* 261(5): 634-645.

Ohki, I., N. Shimotake, N. Fujita, J. Jee, T. Ikegami, M. Nakao and M. Shirakawa (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell.* 105(4): 487-497.

Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* 99(3): 247-257.

Okano, M., S. Xie and E. Li (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet.* 19(3): 219-220.

Okano, M., S. Xie and E. Li (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26(11): 2536-2540.

O'Neill, L. P. and B. M. Turner (2003). Immunoprecipitation of native chromatin: NChIP. *Methods.* 31(1): 76-82.

Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng and T. H. Bestor (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature.* 448(7154): 714-717.

Otani, J., T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi and M. Shirakawa (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.* 10(11): 1235-1241.

Papp, B. and J. Muller (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20(15): 2041-2054.
Pasini, D., P. A. Cloos, J. Walfridsson, L. Olsson, J. P. Bukowski, J. V. Johansen, M. Bak, N. Tommerup, J. Rappsilber and K. Helin (2010a). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature.* 464(7286): 306-310.
Pasini, D., M. Malatesta, H. R. Jung, J. Walfridsson, A. Willer, L. Olsson, J. Skotte, A. Wutz, B. Porse, O. N. Jensen and K. Helin (2010b). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* 38(15): 4958-4969.
Pastor, W. A., U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal and A. Rao (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 473(7347): 394-397.
Patel, K., J. Dickson, S. Din, K. Macleod, D. Jodrell and B. Ramsahoye (2010). Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Res.* 38(13): 4313-4324.
Pauler, F. M., M. A. Sloane, R. Huang, K. Regha, M. V. Koerner, I. Tamir, A. Sommer, A. Aszodi, T. Jenuwein and D. P. Barlow (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* 19(2): 221-233.
Pehrson, J. R. and V. A. Fried (1992). MacroH2A, a core histone containing a large nonhistone region. *Science.* 257(5075): 1398-1400.
Pelizzola, M., Y. Koga, A. E. Urban, M. Krauthammer, S. Weissman, R. Halaban and A. M. Molinaro (2008). MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. *Genome Res.* 18(10): 1652-1659.
Peng, J. C., A. Valouev, T. Swigut, J. Zhang, Y. Zhao, A. Sidow and J. Wysocka (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell.* 139(7): 1290-1302.
Pennings, S., G. Meersseman and E. M. Bradbury (1991). Mobility of positioned nucleosomes on 5 S rDNA. *J Mol Biol.* 220(1): 101-110.
Peterson, R. L., T. Papenbrock, M. M. Davda and A. Awgulewitsch (1994). The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions. *Mech Dev.* 47(3): 253-260.
Phalke, S., O. Nickel, D. Walluscheck, F. Hortig, M. C. Onorati and G. Reuter (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet.* 41(6): 696-702.
Phillips, J. E. and V. G. Corces (2009). CTCF: master weaver of the genome. *Cell.* 137(7): 1194-1211.
Pichler, G., P. Wolf, C. S. Schmidt, D. Meilinger, K. Schneider, C. Frauer, K. Fellinger, A. Rottach and H. Leonhardt (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem.* 112(9): 2585-2593.
Popp, C., W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen and W. Reik (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature.* 463(7284): 1101-1105.
Prendergast, G. C. and E. B. Ziff (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science.* 251(4990): 186-189.
Prokhortchouk, A., B. Hendrich, H. Jorgensen, A. Ruzov, M. Wilm, G. Georgiev, A. Bird and E. Prokhortchouk (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* 15(13): 1613-1618.
Prokhortchouk, A., O. Sansom, J. Selfridge, I. M. Caballero, S. Salozhin, D. Aithozhina, L. Cerchietti, F. G. Meng, L. H. Augenlicht, J. M. Mariadason, B. Hendrich, A. Melnick, E. Prokhortchouk, A. Clarke and A. Bird (2006). Kaiso-deficient mice show resistance to intestinal cancer. *Mol Cell Biol.* 26(1): 199-208.
Qin, W., H. Leonhardt and F. Spada (2011). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem.* 112(2): 439-444.
Quivoron, C., L. Couronne, V. Della Valle, C. K. Lopez, I. Plo, O. Wagner-Ballon, M. Do Cruzeiro, F. Delhommeau, B. Arnulf, M. H. Stern, L. Godley, P. Opolon, H. Tilly, E. Solary, Y.

Duffourd, P. Dessen, H. Merle-Beral, F. Nguyen-Khac, M. Fontenay, W. Vainchenker, C. Bastard, T. Mercher and O. A. Bernard (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell.* 20(1): 25-38.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn and J. Wysocka (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 470(7333): 279-283.

Rai, K., S. Chidester, C. V. Zavala, E. J. Manos, S. R. James, A. R. Karpf, D. A. Jones and B. R. Cairns (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.* 21(3): 261-266.

Rai, K., I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones and B. R. Cairns (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell.* 135(7): 1201-1212.

Rai, K., I. F. Jafri, S. Chidester, S. R. James, A. R. Karpf, B. R. Cairns and D. A. Jones (2010). Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. *J Biol Chem.* 285(6): 4110-4121.

Rakyan, V. K., T. A. Down, S. Maslau, T. Andrew, T. P. Yang, H. Beyan, P. Whittaker, O. T. McCann, S. Finer, A. M. Valdes, R. D. Leslie, P. Deloukas and T. D. Spector (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 20(4): 434-439.

Ramirez-Carrozzi, V. R., D. Braas, D. M. Bhatt, C. S. Cheng, C. Hong, K. R. Doty, J. C. Black, A. Hoffmann, M. Carey and S. T. Smale (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell.* 138(1): 114-128.

Ramsahoye, B. H., D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird and R. Jaenisch (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A.* 97(10): 5237-5242.

Rangasamy, D., L. Berven, P. Ridgway and D. J. Tremethick (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.* 22(7): 1599-1607.

Rank, G., M. Prestel and R. Paro (2002). Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. *Mol Cell Biol.* 22(22): 8026-8034.

Ratnam, S., C. Mertineit, F. Ding, C. Y. Howell, H. J. Clarke, T. H. Bestor, J. R. Chaillet and J. M. Trasler (2002). Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. *Dev Biol.* 245(2): 304-314.

Rauch, T., Z. Wang, X. Zhang, X. Zhong, X. Wu, S. K. Lau, K. H. Kernstine, A. D. Riggs and G. P. Pfeifer (2007). Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A.* 104(13): 5527-5532.

Rauch, T. A., X. Wu, X. Zhong, A. D. Riggs and G. P. Pfeifer (2009). A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A.* 106(3): 671-678.

Rayasam, G. V., O. Wendling, P. O. Angrand, M. Mark, K. Niederreither, L. Song, T. Lerouge, G. L. Hager, P. Chambon and R. Losson (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* 22(12): 3153-3163.

Rebollo, R., M. M. Karimi, M. Bilenky, L. Gagnier, K. Miceli-Royer, Y. Zhang, P. Goyal, T. M. Keane, S. Jones, M. Hirst, M. C. Lorincz and D. L. Mager (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* 7(9): e1002301.

Richly, H., L. Rocha-Viegas, J. D. Ribeiro, S. Demajo, G. Gundem, N. Lopez-Bigas, T. Nakagawa, S. Rospert, T. Ito and L. Di Croce (2010). Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature.* 468(7327): 1124-1128.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet.* 14(1): 9-25.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 129(7): 1311-1323.

Rodriguez, C., J. Borgel, F. Court, G. Cathala, T. Forne and J. Piette (2010). CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem Biophys Res Commun.* 392(2): 129-134.
Rottach, A., C. Frauer, G. Pichler, I. M. Bonapace, F. Spada and H. Leonhardt (2010). The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.* 38(6): 1796-1804.
Rountree, M. R., K. E. Bachman and S. B. Baylin (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet.* 25(3): 269-277.
Rountree, M. R. and E. U. Selker (1997). DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. *Genes Dev.* 11(18): 2383-2395.
Rowe, H. M., J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P. V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D. B. Constam and D. Trono (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature.* 463(7278): 237-240.
Ruzov, A., D. S. Dunican, A. Prokhortchouk, S. Pennings, I. Stancheva, E. Prokhortchouk and R. R. Meehan (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development.* 131(24): 6185-6194.
Ruzov, A., E. Savitskaya, J. A. Hackett, J. P. Reddington, A. Prokhortchouk, M. J. Madej, N. Chekanov, M. Li, D. S. Dunican, E. Prokhortchouk, S. Pennings and R. R. Meehan (2009). The non-methylated DNA-binding function of Kaiso is not required in early Xenopus laevis development. *Development.* 136(5): 729-738.
Sado, T., M. H. Fenner, S. S. Tan, P. Tam, T. Shioda and E. Li (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol.* 225(2): 294-303.
Sado, T., M. Okano, E. Li and H. Sasaki (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development.* 131(5): 975-982.
Sanchez-Elsner, T., D. Gou, E. Kremmer and F. Sauer (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. *Science.* 311(5764): 1118-1123.
Santi, D. V., C. E. Garrett and P. J. Barr (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell.* 33(1): 9-10.
Santoro, R. and I. Grummt (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell.* 8(3): 719-725.
Santos, F., B. Hendrich, W. Reik and W. Dean (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol.* 241(1): 172-182.
Sarraf, S. A. and I. Stancheva (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell.* 15(4): 595-605.
Sasai, N., E. Matsuda, E. Sarashina, Y. Ishida and M. Kawaichi (2005). Identification of a novel BTB-zinc finger transcriptional repressor, CIBZ, that interacts with CtBP corepressor. *Genes Cells.* 10(9): 871-885.
Sasai, N., M. Nakao and P. A. Defossez (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res.* 38(15): 5015-5022.
Sasaki, Y. T., M. Sano, T. Kin, K. Asai and T. Hirose (2007). Coordinated expression of ncRNAs and HOX mRNAs in the human HOXA locus. *Biochem Biophys Res Commun.* 357(3): 724-730.
Sato, N., M. Kondo and K. Arai (2006). The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. *Biochem Biophys Res Commun.* 344(3): 845-851.
Scarsdale, J. N., H. D. Webb, G. D. Ginder and D. C. Williams, Jr. (2011). Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* 39(15): 6741-6752.
Schaefer, M. and F. Lyko (2010b). Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation. *Nat Genet.* 42(11): 920-921; author reply 921.
Schaefer, M., T. Pollex, K. Hanna, F. Tuorto, M. Meusburger, M. Helm and F. Lyko (2010a). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24(15): 1590-1595.
Schaefer, M., J. P. Steringer and F. Lyko (2008). The Drosophila cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS One.* 3(1): e1414.

Schermelleh, L., A. Haemmer, F. Spada, N. Rosing, D. Meilinger, U. Rothbauer, M. C. Cardoso and H. Leonhardt (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* 35(13): 4301-4312.

Scheuermann, J. C., A. G. de Ayala Alonso, K. Oktaba, N. Ly-Hartig, R. K. McGinty, S. Fraterman, M. Wilm, T. W. Muir and J. Muller (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature.* 465(7295): 243-247.

Schmitges, F. W., A. B. Prusty, M. Faty, A. Stutzer, G. M. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, R. D. Bunker, U. Wirth, T. Bouwmeester, A. Bauer, N. Ly-Hartig, K. Zhao, H. Chan, J. Gu, H. Gut, W. Fischle, J. Muller and N. H. Thoma (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell.* 42(3): 330-341.

Schmitt, S., M. Prestel and R. Paro (2005). Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.* 19(6): 697-708.

Schmitz, K. M., C. Mayer, A. Postepska and I. Grummt (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24(20): 2264-2269.

Schorderet, P. and D. Duboule (2011). Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.* 7(5): e1002071.

Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg and T. Jenuwein (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* 18(11): 1251-1262.

Schroeder, D. I., P. Lott, I. Korf and J. M. Lasalle (2011). Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res.*

Schroth, G. P., P. Yau, B. S. Imai, J. M. Gatewood and E. M. Bradbury (1990). A NMR study of mobility in the histone octamer. *FEBS Lett.* 268(1): 117-120.

Schuettengruber, B. and G. Cavalli (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development.* 136(21): 3531-3542.

Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc and G. Cavalli (2007). Genome regulation by polycomb and trithorax proteins. *Cell.* 128(4): 735-745.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek, B. Tolhuis, M. van Lohuizen, A. Tanay and G. Cavalli (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol.* 7(1): e13.

Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon and V. Pirrotta (2010). Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet.* 6(1): e1000805.

Schwartz, Y. B. and V. Pirrotta (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet.* 8(1): 9-22.

Seenundun, S., S. Rampalli, Q. C. Liu, A. Aziz, C. Palii, S. Hong, A. Blais, M. Brand, K. Ge and F. J. Dilworth (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *EMBO J.* 29(8): 1401-1411.

Sen, G. L., J. A. Reuter, D. E. Webster, L. Zhu and P. A. Khavari (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature.* 463(7280): 563-567.

Sessa, L., A. Breiling, G. Lavorgna, L. Silvestri, G. Casari and V. Orlando (2007). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA.* 13(2): 223-239.

Severin, P. M., X. Zou, H. E. Gaub and K. Schulten (2011). Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.*

Sharif, J., M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano and H. Koseki (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature.* 450(7171): 908-912.

Sharma, S., D. D. De Carvalho, S. Jeong, P. A. Jones and G. Liang (2011). Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genet.* 7(2): e1001286.

Sharp, A. J., E. Stathaki, E. Migliavacca, M. Brahmachary, S. B. Montgomery, Y. Dupre and S. E. Antonarakis (2011). DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.*

Shen, L., G. Gao, Y. Zhang, H. Zhang, Z. Ye, S. Huang, J. Huang and J. Kang (2010). A single amino acid substitution confers enhanced methylation activity of mammalian Dnmt3b on chromatin DNA. *Nucleic Acids Res.* 38(18): 6054-6064.
Shen, L., Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland and J. P. Issa (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* 3(10): 2023-2036.
Shen, X., W. Kim, Y. Fujiwara, M. D. Simon, Y. Liu, M. R. Mysliwiec, G. C. Yuan, Y. Lee and S. H. Orkin (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell.* 139(7): 1303-1314.
Shogren-Knaak, M., H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie and C. L. Peterson (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* 311(5762): 844-847.
Simon, J., A. Chiang and W. Bender (1992). Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development.* 114(2): 493-505.
Sing, A., D. Pannell, A. Karaiskakis, K. Sturgeon, M. Djabali, J. Ellis, H. D. Lipshitz and S. P. Cordes (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell.* 138(5): 885-897.
Skene, P. J., R. S. Illingworth, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, R. Andrews and A. P. Bird (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell.* 37(4): 457-468.
Smallwood, A., P. O. Esteve, S. Pradhan and M. Carey (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21(10): 1169-1178.
Smallwood, S. A., S. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S. R. Andrews and G. Kelsey (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet.* 43(8): 811-814.
Song, J., O. Rechkoblit, T. H. Bestor and D. J. Patel (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science.* 331(6020): 1036-1040.
Song, J. J., J. D. Garlick and R. E. Kingston (2008). Structural basis of histone H4 recognition by p55. *Genes Dev.* 22(10): 1313-1318.
Soshnikova, N. and D. Duboule (2009a). Epigenetic regulation of vertebrate Hox genes: a dynamic equilibrium. *Epigenetics.* 4(8): 537-540.
Soshnikova, N. and D. Duboule (2009b). Epigenetic temporal control of mouse Hox genes in vivo. *Science.* 324(5932): 1320-1323.
Soshnikova, N., T. Montavon, M. Leleu, N. Galjart and D. Duboule (2010). Functional analysis of CTCF during mammalian limb development. *Dev Cell.* 19(6): 819-830.
Spada, F., A. Haemmer, D. Kuch, U. Rothbauer, L. Schermelleh, E. Kremmer, T. Carell, G. Langst and H. Leonhardt (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J Cell Biol.* 176(5): 565-571.
Sproul, D., C. Nestor, J. Culley, J. H. Dickson, J. M. Dixon, D. J. Harrison, R. R. Meehan, A. H. Sims and B. H. Ramsahoye (2011). Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A.* 108(11): 4364-4369.
Stancheva, I., O. El-Maarri, J. Walter, A. Niveleau and R. R. Meehan (2002). DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. *Dev Biol.* 243(1): 155-165.
Stancheva, I., C. Hensey and R. R. Meehan (2001). Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *EMBO J.* 20(8): 1963-1973.
Stancheva, I. and R. R. Meehan (2000). Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev.* 14(3): 313-327.
Stein, R., Y. Gruenbaum, Y. Pollack, A. Razin and H. Cedar (1982). Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc Natl Acad Sci U S A.* 79(1): 61-65.
Stock, J. K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A. G. Fisher and A. Pombo (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol.* 9(12): 1428-1435.
Strubbe, G., C. Popp, A. Schmidt, A. Pauli, L. Ringrose, C. Beisel and R. Paro (2011). Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc Natl Acad Sci U S A.* 108(14): 5572-5577.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl Acad Sci U S A.* 82(24): 8419-8423.
Sun, L. Q., D. W. Lee, Q. Zhang, W. Xiao, E. H. Raabe, A. Meeker, D. Miao, D. L. Huso and R. J. Arceci (2004). Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. *Genes Dev.* 18(9): 1035-1046.
Suzuki, M., Q. Jing, D. Lia, M. Pascual, A. McLellan and J. M. Greally (2010). Optimized design and data analysis of tag-based cytosine methylation assays. *Genome Biol.* 11(4): R36.
Suzuki, M. M. and A. Bird (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet.* 9(6): 465-476.
Szulwach, K. E., X. Li, Y. Li, C. X. Song, J. W. Han, S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y. S. Yoon, B. Ren, C. He and P. Jin (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7(6): e1002154.
Tachibana, M., Y. Matsumura, M. Fukuda, H. Kimura and Y. Shinkai (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J.* 27(20): 2681-2690.
Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato and Y. Shinkai (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* 16(14): 1779-1791.
Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 324(5929): 930-935.
Takebayashi, S., T. Tamura, C. Matsuoka and M. Okano (2007). Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. *Mol Cell Biol.* 27(23): 8243-8258.
Takeshita, K., I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa and S. Tajima (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc Natl Acad Sci U S A.* 108(22): 9055-9059.
Tamkun, J. W., R. Deuring, M. P. Scott, M. Kissinger, A. M. Pattatucci, T. C. Kaufman and J. A. Kennison (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell.* 68(3): 561-572.
Tan, M., H. Luo, S. Lee, F. Jin, J. S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocka, Y. Ye, S. Khochbin, B. Ren and Y. Zhao (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell.* 146(6): 1016-1028.
Tanay, A., A. H. O'Donnell, M. Damelin and T. H. Bestor (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc Natl Acad Sci U S A.* 104(13): 5521-5526.
Tao, Y., S. Liu, V. Briones, T. M. Geiman and K. Muegge (2011b). Treatment of breast cancer cells with DNA demethylating agents leads to a release of Pol II stalling at genes with DNA-hypermethylated regions upstream of TSS. *Nucleic Acids Res.*
Tao, Y., S. Xi, V. Briones and K. Muegge (2010). Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation. *PLoS One.* 5(2): e9163.
Tao, Y., S. Xi, J. Shan, A. Maunakea, A. Che, V. Briones, E. Y. Lee, T. Geiman, J. Huang, R. Stephens, R. M. Leighty, K. Zhao and K. Muegge (2011a). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proc Natl Acad Sci U S A.* 108(14): 5626-5631.
Tazi, J. and A. Bird (1990). Alternative chromatin structure at CpG islands. *Cell.* 60(6): 909-920.
Terranova, R., H. Agherbi, A. Boned, S. Meresse and M. Djabali (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A.* 103(17): 6629-6634.
Thoma, F., T. Koller and A. Klug (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol.* 83(2 Pt 1): 403-427.
Thomas, J. O., C. Rees and J. T. Finch (1992). Cooperative binding of the globular domains of histones H1 and H5 to DNA. *Nucleic Acids Res.* 20(2): 187-194.

Thompson, R. F., M. Reimers, B. Khulan, M. Gissot, T. A. Richmond, Q. Chen, X. Zheng, K. Kim and J. M. Greally (2008). An analytical pipeline for genomic representations used for cytosine methylation studies. *Bioinformatics.* 24(9): 1161-1167.

Thomson, J. P., P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth and A. Bird (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature.* 464(7291): 1082-1086.

Tie, F., R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik, A. Zlobin, M. O. Diaz, P. C. Scacheri and P. J. Harte (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development.* 136(18): 3131-3141.

Tolhuis, B., M. Blom, R. M. Kerkhoven, L. Pagie, H. Teunissen, M. Nieuwland, M. Simonis, W. de Laat, M. van Lohuizen and B. van Steensel (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* 7(3): e1001343.

Tomizawa, S., H. Kobayashi, T. Watanabe, S. Andrews, K. Hata, G. Kelsey and H. Sasaki (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development.* 138(5): 811-820.

Trowbridge, J. J., J. W. Snow, J. Kim and S. H. Orkin (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell.* 5(4): 442-449.

Tsumura, A., T. Hayakawa, Y. Kumaki, S. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. Nakayama and M. Okano (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells.* 11(7): 805-814.

Tsutsui, K. M., K. Sano, O. Hosoya, T. Miyamoto and K. Tsutsui (2011). Nuclear protein LEDGF/p75 recognizes supercoiled DNA by a novel DNA-binding domain. *Nucleic Acids Res.* 39(12): 5067-5081.

Turek-Plewa, J. and P. P. Jagodzinski (2005). The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett.* 10(4): 631-647.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays.* 22(9): 836-845.

Unoki, M., T. Nishidate and Y. Nakamura (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene.* 23(46): 7601-7610.

Valdes-Mora, F., J. Z. Song, A. L. Statham, D. Strbenac, M. D. Robinson, S. S. Nair, K. I. Patterson, D. J. Tremethick, C. Stirzaker and S. J. Clark (2011). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res.*

Valinluck, V. and L. C. Sowers (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 67(3): 946-950.

van der Lugt, N. M., J. Domen, K. Linders, M. van Roon, E. Robanus-Maandag, H. te Riele, M. van der Valk, J. Deschamps, M. Sofroniew, M. van Lohuizen and et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev.* 8(7): 757-769.

van der Vlag, J. and A. P. Otte (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet.* 23(4): 474-478.

Van Emburgh, B. O. and K. D. Robertson (2011). Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Res.* 39(12): 4984-5002.

van Holde, K. and J. Zlatanova (1999). The nucleosome core particle: does it have structural and physiologic relevance? *Bioessays.* 21(9): 776-780.

Vastenhouw, N. L., Y. Zhang, I. G. Woods, F. Imam, A. Regev, X. S. Liu, J. Rinn and A. F. Schier (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature.* 464(7290): 922-926.

Velasco, G., F. Hube, J. Rollin, D. Neuillet, C. Philippe, H. Bouzinba-Segard, A. Galvani, E. Viegas-Pequignot and C. Francastel (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A.* 107(20): 9281-9286.

Vernimmen, D., M. D. Lynch, M. De Gobbi, D. Garrick, J. A. Sharpe, J. A. Sloane-Stanley, A. J. Smith and D. R. Higgs (2011). Polycomb eviction as a new distant enhancer function. *Genes Dev.* 25(15): 1583-1588.

Villa, R., L. Morey, V. A. Raker, M. Buschbeck, A. Gutierrez, F. De Santis, M. Corsaro, F. Varas, D. Bossi, S. Minucci, P. G. Pelicci and L. Di Croce (2006). The methyl-CpG binding protein MBD1 is required for PML-RARalpha function. *Proc Natl Acad Sci U S A.* 103(5): 1400-1405.

Vire, E., C. Brenner, R. Deplus, L. Blanchon, M. Fraga, C. Didelot, L. Morey, A. Van Eynde, D. Bernard, J. M. Vanderwinden, M. Bollen, M. Esteller, L. Di Croce, Y. de Launoit and F. Fuks (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature.* 439(7078): 871-874.

Voncken, J. W., B. A. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps and M. van Lohuizen (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A.* 100(5): 2468-2473.

Voo, K. S., D. L. Carlone, B. M. Jacobsen, A. Flodin and D. G. Skalnik (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol.* 20(6): 2108-2121.

Walsh, C. P., J. R. Chaillet and T. H. Bestor (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet.* 20(2): 116-117.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones and Y. Zhang (2004a). Role of histone H2A ubiquitination in Polycomb silencing. *Nature.* 431(7010): 873-878.

Wang, J., S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li and T. Chen (2009a). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet.* 41(1): 125-129.

Wang, J. and G. M. Shackleford (1996). Murine Wnt10a and Wnt10b: cloning and expression in developing limbs, face and skin of embryos and in adults. *Oncogene.* 13(7): 1537-1544.

Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms and H. Y. Chang (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature.* 472(7341): 120-124.

Wang, L., J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis and R. S. Jones (2004b). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell.* 14(5): 637-646.

Wang, P., C. Lin, E. R. Smith, H. Guo, B. W. Sanderson, M. Wu, M. Gogol, T. Alexander, C. Seidel, L. M. Wiedemann, K. Ge, R. Krumlauf and A. Shilatifard (2009c). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol.* 29(22): 6074-6085.

Wang, Z., C. Zang, K. Cui, D. E. Schones, A. Barski, W. Peng and K. Zhao (2009b). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell.* 138(5): 1019-1031.

Watt, F. and P. L. Molloy (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* 2(9): 1136-1143.

Weber, A., J. Marquardt, D. Elzi, N. Forster, S. Starke, A. Glaum, D. Yamada, P. A. Defossez, J. Delrow, R. N. Eisenman, H. Christiansen and M. Eilers (2008). Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. *EMBO J.* 27(11): 1563-1574.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet.* 39(4): 457-466.

Weinhofer, I., E. Hehenberger, P. Roszak, L. Hennig and C. Kohler (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet.* 6(10).

Whitehouse, I., A. Flaus, B. R. Cairns, M. F. White, J. L. Workman and T. Owen-Hughes (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature.* 400(6746): 784-787.

Wiench, M., S. John, S. Baek, T. A. Johnson, M. H. Sung, T. Escobar, C. A. Simmons, K. H. Pearce, S. C. Biddie, P. J. Sabo, R. E. Thurman, J. A. Stamatoyannopoulos and G. L. Hager (2011). DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* 30(15): 3028-3039.

Wienholz, B. L., M. S. Kareta, A. H. Moarefi, C. A. Gordon, P. A. Ginno and F. Chedin (2010). DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS Genet.* 6(9).

Williams, K., J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber and K. Helin (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature.* 473(7347): 343-348.

Woo, C. J., P. V. Kharchenko, L. Daheron, P. J. Park and R. E. Kingston (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell.* 140(1): 99-110.

Woodcock, C. L., A. I. Skoultchi and Y. Fan (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* 14(1): 17-25.

Wossidlo, M., T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik and J. Walter (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun.* 2: 241.

Wu, H., V. Coskun, J. Tao, W. Xie, W. Ge, K. Yoshikawa, E. Li, Y. Zhang and Y. E. Sun (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science.* 329(5990): 444-448.

Wu, H., A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun and Y. Zhang (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* 25(7): 679-684.

Wu, X., Y. Gong, J. Yue, B. Qiang, J. Yuan and X. Peng (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. *Nucleic Acids Res.* 36(11): 3590-3599.

Xi, H., Y. Yu, Y. Fu, J. Foley, A. Halees and Z. Weng (2007b). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res.* 17(6): 798-806.

Xi, S., H. Zhu, H. Xu, A. Schmidtmann, T. M. Geiman and K. Muegge (2007a). Lsh controls Hox gene silencing during development. *Proc Natl Acad Sci U S A.* 104(36): 14366-14371.

Xie, S., Z. Wang, M. Okano, M. Nogami, Y. Li, W. W. He, K. Okumura and E. Li (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene.* 236(1): 87-95.

Xu, C., C. Bian, R. Lam, A. Dong and J. Min (2011a). The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat Commun.* 2: 227.

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo and E. Viegas-Pequignot (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature.* 402(6758): 187-191.

Xu, Y., F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang and Y. G. Shi (2011b). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell.* 42(4): 451-464.

Yan, Q., E. Cho, S. Lockett and K. Muegge (2003a). Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. *Mol Cell Biol.* 23(23): 8416-8428.

Yan, Q., J. Huang, T. Fan, H. Zhu and K. Muegge (2003b). Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. *EMBO J.* 22(19): 5154-5162.

Yan, X. J., J. Xu, Z. H. Gu, C. M. Pan, G. Lu, Y. Shen, J. Y. Shi, Y. M. Zhu, L. Tang, X. W. Zhang, W. X. Liang, J. Q. Mi, H. D. Song, K. Q. Li, Z. Chen and S. J. Chen (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet.* 43(4): 309-315.

Yoder, J. A. and T. H. Bestor (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet.* 7(2): 279-284.

Yoder, J. A., N. S. Soman, G. L. Verdine and T. H. Bestor (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol.* 270(3): 385-395.

Yoon, H. G., D. W. Chan, A. B. Reynolds, J. Qin and J. Wong (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell.* 12(3): 723-734.

Yuan, W., M. Xu, C. Huang, N. Liu, S. Chen and B. Zhu (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem.* 286(10): 7983-7989.

Yuzyuk, T., T. H. Fakhouri, J. Kiefer and S. E. Mango (2009). The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in C. elegans embryos. *Dev Cell.* 16(5): 699-710.

Zemach, A., I. E. McDaniel, P. Silva and D. Zilberman (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 328(5980): 916-919.

Zemach, A. and D. Zilberman (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol.* 20(17): R780-785.

Zhang, H., B. Niu, J. F. Hu, S. Ge, H. Wang, T. Li, J. Ling, B. N. Steelman, G. Qian and A. R. Hoffman (2011b). Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. *J Cell Biol.* 193(3): 475-487.

Zhang, H., D. N. Roberts and B. R. Cairns (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell.* 123(2): 219-231.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle and A. Jeltsch (2010b). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* 38(13): 4246-4253.

Zhang, Y., H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird and D. Reinberg (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 13(15): 1924-1935.

Zhang, Y., C. Rohde, S. Tierling, T. P. Jurkowski, C. Bock, D. Santacruz, S. Ragozin, R. Reinhardt, M. Groth, J. Walter and A. Jeltsch (2009). DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet.* 5(3): e1000438.

Zhang, Z., A. Jones, C. W. Sun, C. Li, C. W. Chang, H. Y. Joo, Q. Dai, M. R. Mysliwiec, L. C. Wu, Y. Guo, W. Yang, K. Liu, K. M. Pawlik, H. Erdjument-Bromage, P. Tempst, Y. Lee, J. Min, T. M. Townes and H. Wang (2010a). PRC2 Complexes with JARID2, and esPRC2p48 in ES Cells to Modulate ES Cell Pluripotency and Somatic Cell Reprogramming. *Stem Cells.*

Zhang, Z., C. J. Wippo, M. Wal, E. Ward, P. Korber and B. F. Pugh (2011a). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science.* 332(6032): 977-980.

Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky and J. T. Lee (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell.* 40(6): 939-953.

Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 322(5902): 750-756.

Zhao, Q., G. Rank, Y. T. Tan, H. Li, R. L. Moritz, R. J. Simpson, L. Cerruti, D. J. Curtis, D. J. Patel, C. D. Allis, J. M. Cunningham and S. M. Jane (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol.* 16(3): 304-311.

Zhao, X., T. Ueba, B. R. Christie, B. Barkho, M. J. McConnell, K. Nakashima, E. S. Lein, B. D. Eadie, A. R. Willhoite, A. R. Muotri, R. G. Summers, J. Chun, K. F. Lee and F. H. Gage (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc Natl Acad Sci U S A.* 100(11): 6777-6782.

Zhu, H., T. M. Geiman, S. Xi, Q. Jiang, A. Schmidtmann, T. Chen, E. Li and K. Muegge (2006). Lsh is involved in de novo methylation of DNA. *EMBO J.* 25(2): 335-345.

Zilberman, D., D. Coleman-Derr, T. Ballinger and S. Henikoff (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature.* 456(7218): 125-129.

Zilberman, D., M. Gehring, R. K. Tran, T. Ballinger and S. Henikoff (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet.* 39(1): 61-69.